Waseda University

Department of Mathematics

Master Thesis in Mathematical Statistics

Nonparametric Methods in Time Series Analysis

Author:
Yan Liu

Supervisor:
Dr. Masanobu Taniguchi

April 12, 2014
Academic Year 2012/2013

Contents

Preface

Nonparametric methods have been developed for statistical analysis of univariate and multivariate observations in the area of time series analysis to carry out the problem of hypothesis testing and inference in these two decades. Compared with parametric methods, the notable feature is that they have fewer assumptions on the model or the distribution of its innovations. Parametric statistics is still, however, the mainstream in the mathematical statistics according to its obviousness and tractability. The theory of parametric case has been established since Le Cam proposed his three lemmas and the sophisticated logic is attractable. We treasure the theory in the Appendix in this Thesis. Nevertheless, the assumptions of the parameters restrict its further development and we start the research on the nonparametric methods.

Recently, two of the most remarkable nonparametric methods are empirical likelihood ratio method and rank-based method. The former method was introduced by Owen (1988), who applied it to the i.i.d data and showed the validity of the empirical likelihood ratio statistic. For dependent data, Monti (1997), Ogata and Taniguchi (2010) showed that the asymptotic distribution of the empirical likelihood ratio statistic is asymptotically χ^2-distributed based on the Whittle likelihood type estimating function under the regular conditions. As a result, we can construct confidence interval for pivotal quantity such as the coefficients in a predictor and autocorrelation coefficients in multivariate stationary processes, etc. The latter method is well known for Spearman's rank correlation. Its history is so long that we can not give accurate literature for it. The story of i.i.d case is well established in the book by Hájek (1968). Hallin et al. (1985, 1987) and Hallin and Puri (1991) extended the case to the depend case. Garel and Hallin (1995) also proposed the LAN theorem for multiple time series and extended the result further to the semiparametric case.

In the last few decades, it is founded in the most cases that the behavior of data is neither independent nor Gaussian. Furthermore, heavy-tailed data have been observed in a variety of fields involving electrical engineering, hydrology, finance and physical systems (See Nolan (2012) and Samoradnitsky and Taqqu (2000)). In particular, Fama (1965) and Mandelbrot (1963) gave the economic and financial examples that showed such data are poorly represented by Gaussian model. A good choice for model in this case is proposed to use innovations with regularly varying tail with index α, which is a necessary and sufficient condition for the limit to be stable random variables. However, stable random variables are defined by its Fourier transformation, and there is no analytic representation for the probability density function. This is a reason why the model driven by stable innovations are difficult to analyze by maximum likelihood method or in a parametric way.

In this research, we apply the nonparametric methods to the stable case and investigated the property of the methods. In the empirical likelihood ratio case, we applied the self-normalized method to the generalized linear case and obtained the asymptotic distribution of the statistics. In the sequent chapter, we probed into the self-normalized random variables' asymptotic moments. In the last chapter, we follow the idea introduced by Prof. Hallin and summarized his result.

Chapter 1

Empirical Likelihood Method

1.1 Introduction

Recently, the multivariate data with infinite variance appear in various fields like finance, economics and hydrology. To model these phenomena, one choice is to apply generalized linear process to the case.

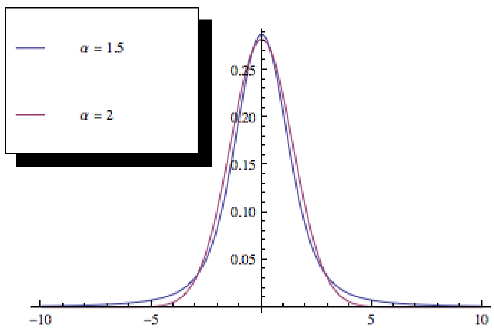


Figure 1.1: Stable distribution ($\alpha = 1.5$) and Normal distribution ($\alpha = 2$)

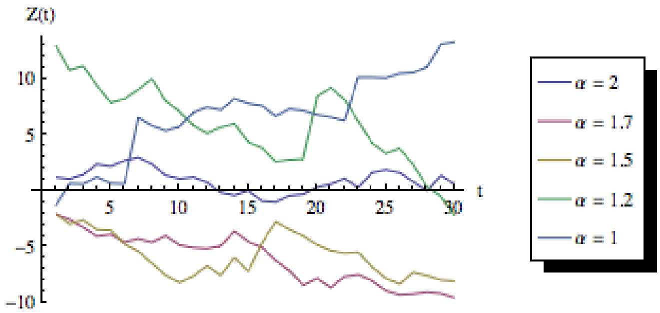


Figure 1.2: AR(1) model driven by Stable distribution ($\alpha = 2, 1.7, 1.5, 1.2, 1$)

Figure 1.1 shows the probability density of both stable distribution and normal distribution ($\alpha = 1.5$ and $\alpha = 2$ with $\sigma = 1$). Figure 1.2 shows a linear process driven by the stable innovation ($\alpha = 2, 1.7, 1.5, 1.2, 1$) with $\sigma = 1$. Although the difference between the probability density function is not seemingly so much, it is easy to see that the linear process with smaller index waves more dynamically in its range. In fact, only the case of $\alpha = 2$ has finite variance, and the others do not. Accordingly, a process under regular conditions is much more different from the stable one.

For 1-dimensional linear process with infinite variance innovations, Davis and Resnick (1985a, 1985b and 1986) investigated the sample autocorrelation function (ACF) at lag h, and derived the consistency of ACF. Resnick and Stărică (1998) gave a consistent estimator of the tail index α. In view of the frequency domain approach, Klüppelberg and Mikosch (1993, 1994 and 1996) proposed self-normalized periodogram because the expectation of the usual periodogram does not exist, and introduced some methods for parameter estimation and hypothesis testing. Then, they showed that for any frequencies, self-normalized periodogram converges to a random variable with finite second moment, and proved the convergence of the integral functional of the self-normalized periodogram.

In this paper, we apply nonparametrical method to the discrete d-dimensional linear process

$$\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \Psi(j)\boldsymbol{Z}(t-j). \tag{1.1.1}$$

It is natural to express the process nonparametrically partly because finite parametric models often can not describe real data sufficiently, and partly because there is no general solution of probability density function for stable distribution. Recently economists and quantitive analysts have introduced stable stochastic models to asset returns in econometrics and finance. In such situations, what we are interested in is to test statistical hypothesis on the pivotal quantity "$\boldsymbol{\theta} = \boldsymbol{\theta}_0$", such as the correlation between the different realizations. To achieve this goal, Monti (1997) and Ogata and Taniguchi (2010) employed the empirical likelihood to construct confidence sets for linear processes when innovations have finite variance. A plausible way to define the important index $\boldsymbol{\theta}_0$ is Whittle's approach, that is, $\boldsymbol{\theta}_0$ minimizes the disparity

$$D(\boldsymbol{f}_{\boldsymbol{\theta}}, \boldsymbol{g}) = \int_{-\pi}^{\pi} \mathrm{tr}\{\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\boldsymbol{g}(\omega))\}d\omega. \tag{1.1.2}$$

The empirical likelihood ratio function for the problem of testing H: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is defined as

$$R(\boldsymbol{\theta}) = \max_{\omega_1,\dots,\omega_n}\left\{\prod_{t=1}^{n} n\omega_t\,;\ \sum_{t=1}^{n}\omega_t\,\boldsymbol{m}(\lambda_t;\boldsymbol{\theta}) = \boldsymbol{0}, \sum_{t=1}^{n}\omega_t = 1, 0 \le \omega_t \le 1\ ,\ \forall t\right\}, \tag{1.1.3}$$

and then the estimating function takes the form

$$\boldsymbol{m}(\lambda_t;\boldsymbol{\theta}) \equiv \frac{\partial}{\partial\boldsymbol{\theta}}\mathrm{tr}\{\boldsymbol{f}(\lambda_t;\boldsymbol{\theta})^{-1}\boldsymbol{I}_{n,X}(\lambda_t)\}, \quad \lambda_t = \frac{2\pi t}{n} \in (-\pi,\pi]. \tag{1.1.4}$$

For our general stable linear process, we derive the limit distribution of $R(\boldsymbol{\theta}_0)$ with its normalizing factor and construct the confidence interval through a numerical method.

Here it should be noted that our extension to the stable case from the finite variance case requires new asymptotic methods, and we report new aspects of the asymptotics, which are different from the usual ones. Furthermore, we extend the results to the multivariate one, which is extremely important from a viewpoint of practical use. The way to derive the asymptotics of the multivariate case has also new aspects. One is that the self-normalizing factor is not the square root matrix but the norm of the stable series. The other is that we need stronger assumption on the pivotal quantity to have the asymptotic distribution hold.

1.2 Vector α-stable Processes

We consider a d-dimensional vector-valued linear process $\{\boldsymbol{X}(t); t \in \mathbb{Z}\}$ generated by

$$\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \Psi(j)\boldsymbol{Z}(t-j) \tag{1.2.1}$$

where $\Psi(j)$ are $d \times d$ real matrices, and $\{\boldsymbol{Z}(t)\}$ is an independently and identically distributed sequence of symmetric α-stable random vectors whose elements are also independent.

1.2.1 Notations and Preliminaries

First, we give a brief review on the notation in this chapter. we use Bold letters to represent vectors or matrices. For an element in vectors or matrices, we use underscript. For instance, A_j denotes the jth entry in the vector $\boldsymbol{A}$, where A_{ij} denotes the element lying in the ith row and jth column of the matrix $\boldsymbol{A}$.

The argument about the frequency domain makes us unable to escape from the complex numbers. We generally use $\bar{A}$ to denote the complex conjugate of A regardless of whether A is a complex number or a complex matrix.

Next, the usage of characters are as follows. Note that we use $\omega \in [-\pi, \pi]$ for continuous case, and $\lambda_t = \frac{2\pi t}{n} \in (-\pi, \pi]$ for discrete case. For any random vector $\boldsymbol{A}$, the sample autocovariance and the periodogram matrices are defined as

$$\hat{\Gamma}_{n,A}(h) = n^{-2/\alpha} \sum_{t=1}^{n-|h|} \boldsymbol{A}(t)\boldsymbol{A}(t+h)',$$

$$\boldsymbol{I}_{n,A}(\omega) = d_{n,A}(\omega)d_{n,A}(\omega)^*, \quad d_{n,A}(\omega) = n^{-1/\alpha} \sum_{t=1}^{n} \boldsymbol{A}(t)e^{\boldsymbol{i}\omega t},$$

respectively.

There are two norms used in this paper. One is the Euclidean norm, which is denoted by $\|\cdot\|_E$. Secondly, based on the observed stretch $\{\boldsymbol{X}(t), 1 \leq t \leq n\}$, the self-normalized term, denoted by $\|Z\|_N$, is defined as follows:

$$\|Z\|_N \equiv \sqrt{\sum_{t=1}^{n}\sum_{i=1}^{d} Z(t)_i^2}. \tag{1.2.2}$$

It is well known that $Z(t)_i^2$ is in the domain of attraction of a stable limit with $\alpha/2$, and the linear transformation of stable distribution with nonrandom scale is also stable with the same characteristic exponent. Thus the sum $\sum_{i=1}^{d} Z(t)_i^2$ is also in the domain of attraction of a stable limit with $\alpha/2$. The normalized form of vectors is written as

$$\tilde{Z}(t)_i = \frac{Z(t)_i}{\|Z\|_N}, \quad i = 1, \ldots, d. \tag{1.2.3}$$

For the model (1.1.1), we define the true power transfer function $\boldsymbol{g}(\omega)$ by

$$\boldsymbol{g}(\omega) = \Psi(\omega)\Psi(\omega)^*,$$

where $\Psi(\omega) = \sum_{j=0}^{\infty} \Psi(j)e^{\boldsymbol{i}j\omega}$. To derive the asymptotics of the empirical likelihood ratio function, we write a fitted spectral in the parametric way, that is, we use $\boldsymbol{f}(\omega; \boldsymbol{\theta})$, which satisfies assumptions in the next section. It is not necessary that the true spectral density be in the family of fitted spectral densities. Here we define the pivotal value $\boldsymbol{\theta}_0$, which satisfies

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \mathrm{tr}\left[\, \{\boldsymbol{f}(\omega; \boldsymbol{\theta})\}^{-1} \boldsymbol{g}(\omega) \,\right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \boldsymbol{0}. \tag{1.2.4}$$

For example, we can use the constrained family to derive the autocorrelation, the interpolation or the prediction. The empirical likelihood ratio is defined as

$$R(\boldsymbol{\theta}) = \max\{\prod_{t=1}^{n} np_t \Big| \sum_{t=1}^{n} p_t\, \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) = \boldsymbol{0}, \quad p_t \geq 0, \quad \sum_{t=1}^{n} p_t = 1\}, \tag{1.2.5}$$

with any estimating function $\boldsymbol{m}(\lambda_t; \boldsymbol{\theta})$. In the time series literature, the most common estimating function is

$$\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\{\boldsymbol{f}(\lambda_t; \boldsymbol{\theta})^{-1} \boldsymbol{I}_{n,X}(\lambda_t)\}, \tag{1.2.6}$$

which is called the Whittle likelihood. Also, for the brevity, we define moment functions of the estimating function $\boldsymbol{P}_n(\boldsymbol{\theta})$ and $\boldsymbol{S}_n(\boldsymbol{\theta})$ as follows:

$$\boldsymbol{P}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) \tag{1.2.7}$$

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) \boldsymbol{m}(\lambda_t; \boldsymbol{\theta})'. \tag{1.2.8}$$

Assumptions through this chapter are given here:

Assumption 1.2.1. Assume that $\boldsymbol{X}(t)$ is generated by (1.1.1) where

$$\sum_{j=0}^{\infty} j|\Psi(j)_{kl}|^{\delta} < \infty, \quad \text{for } k,l = 1,2,\dots,d \quad \text{where } \delta < \min\{\alpha, 1\}. \tag{1.2.9}$$

Remark 1.2.2. Under this assumption, (1.1.1) is well-defined. See Brockwell and davis (1991) or Petorov (1975).

Define the family $\mathcal{F}$ of the fitted power transfer function as

$$\mathcal{F} = \{\boldsymbol{f}(\omega;\boldsymbol{\theta}) \Big| \boldsymbol{f}(\omega;\boldsymbol{\theta}) = \Big(\sum_{j=0}^{\infty} \Xi(j;\boldsymbol{\theta})e^{\boldsymbol{i}j\omega}\Big)\Big(\sum_{j=0}^{\infty} \Xi(j;\boldsymbol{\theta})e^{\boldsymbol{i}j\omega}\Big)^{*}, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{q}\}, \tag{1.2.10}$$

where

$$\sum_{j=0}^{\infty} j|\Xi(j)_{kl}|^{\delta} < \infty, \quad \text{for } k,l = 1,2,\dots,d \quad \text{where } \delta < \min\{\alpha, 1\}. \tag{1.2.11}$$

Assumption 1.2.3.

(i) Θ is a compact subset of $\mathbb{R}^q$.

(ii) There exists a unique $\boldsymbol{\theta}_0 \in \Theta$ satisfying (1.2.4).

(iii) $\boldsymbol{f}(\omega;\boldsymbol{\theta}) \in \mathcal{F}$ is continuously differentiable with respect to $\boldsymbol{\theta}$.

The assumption below guarantees the convergence of the functional of periodogram by inequality of an application of Theorem 3.1 in Rosinski and Woyczynski (1987).

Assumption 1.2.4. For some $\mu \in (0,\alpha)$ and all $k = 1,\cdots,q$,

$$\sum_{t=1}^{\infty} \Big\| \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_k} \Psi(\omega)^{*} \boldsymbol{f}(\omega;\boldsymbol{\theta})\Psi(\omega)e^{\boldsymbol{i}t\omega} d\omega \Big\|_{E}^{\mu} < \infty. \tag{1.2.12}$$

1.3 Main Result

First the limit of functional form of periodogram is shown in the following theorem, which is a generalization of 1-dimensional result.

Theorem 1.3.1. *Let $(\boldsymbol{X}(t))_{t\in\mathbb{Z}}$ be a linear process as defined in* (1.1.1) *with coefficient matrices $(\Psi(j))_{j\in\mathbb{Z}}$ satisfying* (1.2.9) *and suppose that $\alpha \in (0,2)$. Furthermore, let $\phi_k(\omega), j = 1,\ldots,d$, be $d\times d$ matrix-valued 2π-periodic continuous function with $\phi_k(\omega) = \phi_k(\omega)^*$ such that the Fourier coefficients of $\Psi(\cdot)\phi_k(\cdot)\Psi(\cdot)^*$ are absolutely summable and*

$$\sum_{t=1}^{\infty}\Big\|\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_k}\Psi(\omega)^*\phi_k(\omega)\Psi(\omega)e^{it\omega}d\omega\Big\|_E^{\mu} < \infty \tag{1.3.1}$$

for some $\mu\in(0,\alpha)$ and all $k = 1,\cdots,q$. Then

$$\begin{aligned}(n^{-2/\alpha}\|Z\|_N^2, x_n\int_{-\pi}^{\pi}\mathrm{tr}\Big[\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^*\}\phi_k(\omega)\Big]\,d\omega)\\ \xrightarrow{\mathcal{L}} (\,S_{\alpha/2}, \sum_{i,j=1}^{d}\sum_{h=1}^{\infty}S(h)_{ij}\int_{-\pi}^{\pi}(A(\omega)+\overline{A(\omega)})_{ij}\,d\omega\,),\end{aligned} \tag{1.3.2}$$

where

$$A(\omega) = \Psi(\omega)^*\phi_k(\omega)\Psi(\omega)e^{ih\omega},$$

and $S(h)_{ij}$ is the (i,j)-component of the limit stable random matrix $S(h)$, where

$$x_n\,\hat{\Gamma}_{n,Z}(h) \Rightarrow S(h) \quad \textit{for } h = 1,2,\ldots.$$

Theorem 1.3.2. *Let $(\boldsymbol{X}(t))_{t\in\mathbb{Z}}$ be a linear process as defined in* (1.1.1) *with coefficient matrices $(\Psi(j))_{j\in\mathbb{Z}}$ satisfying* (1.2.9) *and suppose that $\alpha\in[1,2)$. Under Assumptions* 1.2.3 *and* 1.2.4*, if*

$$\frac{\partial}{\partial\boldsymbol{\theta}}\int_{-\pi}^{\pi}\Psi(\omega)^*\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\Psi(\omega)\,d\omega\,\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \boldsymbol{0}, \tag{1.3.3}$$

we have

$$-2\frac{x_n^2}{n}\log R(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \boldsymbol{V}'\boldsymbol{W}^{-1}\boldsymbol{V} \quad \textit{under } H\textit{: } \boldsymbol{\theta} = \boldsymbol{\theta}_0,$$

where

$$\boldsymbol{V} = \frac{1}{2\pi}\sum_{i,j=1}^{d}\sum_{h=1}^{\infty}\frac{S(h)_{ij}}{S_{\alpha/2}}\begin{pmatrix}\int_{-\pi}^{\pi}(B_1(\omega)+\overline{B_1(\omega)})_{ij}d\omega\\ \int_{-\pi}^{\pi}(B_2(\omega)+\overline{B_2(\omega)})_{ij}d\omega\\ \vdots\\ \int_{-\pi}^{\pi}(B_q(\omega)+\overline{B_q(\omega)})_{ij}d\omega,\end{pmatrix} \tag{1.3.4}$$

with

$$B_k(\omega) = \Psi(\omega)^*\frac{\partial}{\partial\theta_k}\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\Psi(\omega) \quad k = 1,\ldots,q,$$

and the component of $\boldsymbol{W}$ is expressed as

$$W_{ab} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\right.$$
$$\left.+\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\right)d\omega, \quad (1.3.5)$$

where $\tilde{\boldsymbol{g}}(\omega)$ is defined as

$$\tilde{\boldsymbol{g}}(\omega) = \Psi(\omega)\Sigma_{\tilde{Z}}\Psi(\omega)^*. \quad (1.3.6)$$

Corollary 1.3.3. With the same assumptions and condition (1.3.3), we have

$$-2\frac{x_n^2}{n}\log R(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \boldsymbol{V}'\boldsymbol{W}^{-1}\boldsymbol{V} \quad \text{under } H\text{: } \boldsymbol{\theta} = \boldsymbol{\theta}_0,$$

where V is defined above and the component of $\boldsymbol{W}$ can be expressed as

$$W_{ab} = \frac{1}{2\pi d^2}\int_{-\pi}^{\pi}\left(\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\right.$$
$$\left.+\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\right)d\omega. \quad (1.3.7)$$

1.3.1 Numerical Study

We consider the 2-dimemsional VAR(1) model:

$$\boldsymbol{X}(t) + A\boldsymbol{X}(t-1) = \boldsymbol{Z}(\boldsymbol{t}), \quad (1.3.8)$$

where the marginal distributions of $\{\boldsymbol{Z}(t)\}$ are assumed to be i.i.d. symmetric α-stable variable with scale 1 for simplicity. The true coefficient matrix is given by

$$A = \begin{pmatrix} 0.7 & \theta_0 \\ 0.1 & 0.5 \end{pmatrix}.$$

Like other methods, the parameter estimation is possible. First, we define the fitted power transfer function $\boldsymbol{f}(\omega;\boldsymbol{\theta})$ corresponding to the estimating function as

$$\boldsymbol{f}(\omega;\boldsymbol{\theta}) = (I - Be^{i\omega})^{-1}(I - Be^{i\omega})^{-1*}, \quad \text{where} \quad B = \begin{pmatrix} 0.7 & \theta \\ 0.1 & 0.5 \end{pmatrix}.$$

The numerical results in this case are given in Table 1.1.

Table 1.1: 95% confidence intervals (and length) for true parameter. Sample size is 500 and $\alpha = 1.7$.

	θ_0	E.L		(length)
case 1.	0	-0.1242	0.0692	(0.1934)
case 2.	0.5	0.4475	0.5425	(0.0950)
case 3.	0.6	0.0940	0.7607	(0.6667)

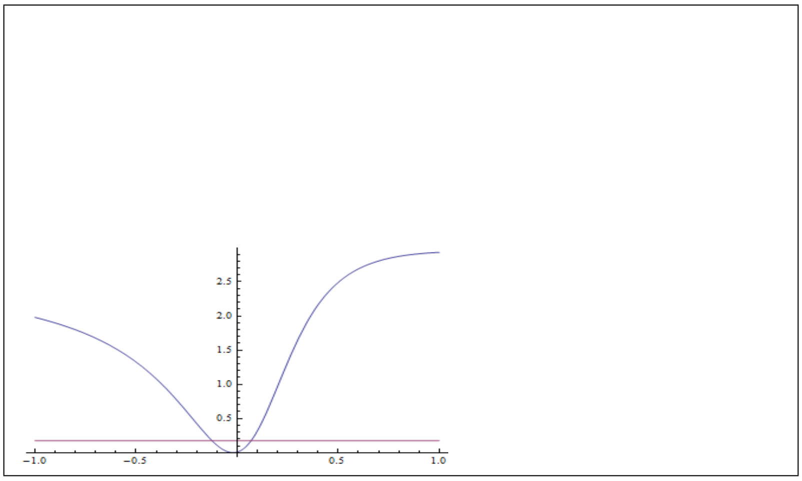


Figure 1.3: The confidence interval in case 1

Next, we examine the (1,1)-component of autocorrelation (See Brockwell and Davis [1991]), which is defined as

$$\rho_{11}(l) = \gamma_{11}(l)/\gamma_{11}(0), \quad l = 0, 1, \dots. \tag{1.3.9}$$

The estimation of this quantity is equivalent to fit the power transfer function

$$(I - Be^{-il\omega})(I - Be^{il\omega})'^{-1}, \tag{1.3.10}$$

where B has the form

$$\begin{pmatrix} \theta & b \\ 0 & c \end{pmatrix}, \quad \text{where } b \text{ and } c \text{ are any constants. (See Appendix in this Chapter.)}$$

The numerical results are given in Table 1.2.

Table 1.2: 95% confidence intervals (and length) for true parameter. Sample size is 2000 and $\alpha = 1.7$.

	l	θ_0	E.L		(length)
case 4.	2	0.700	-0.3987	0.8037	(1.2024)
case 5.	3	0.590	-0.1299	0.8250	(0.9549)
case 6.	4	0.498	-0.5969	0.6759	(1.2728)

Another appealing example is to consider whether the wave structures of the spectra between all components are "close" to each other or not. For simplicity, we formulate this idea in 2-dimensional case and assume the true power transfer function $\boldsymbol{g}(\omega)$ is

$$\boldsymbol{g}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \tilde{\boldsymbol{R}}(k) e^{-ik\omega}. \tag{1.3.11}$$

Then the null hypothesis can be written as

$$H: \quad \tilde{\boldsymbol{R}}(k) = \theta_0 \tilde{\boldsymbol{R}}(j) \quad \text{or} \quad \tilde{\boldsymbol{R}}(k) = \theta_0 \tilde{\boldsymbol{R}}(j)' \quad \text{for some } k \text{ and } j.$$

To test this hypothesis, we set the estimating function $\boldsymbol{m}(\lambda_t;\boldsymbol{\theta})$ with an inverse correlation function $\boldsymbol{f}(\lambda_t;\boldsymbol{\theta})^{-1}$, which was first introduced in Cleveland (1972), and deeply discussed by Bhansali (1980). Let

$$\boldsymbol{f}(\omega;\theta)^{-1} = (e^{k\omega} + e^{-k\omega}) \begin{pmatrix} \theta & 0 \\ 0 & \theta \end{pmatrix} + (e^{j\omega} + e^{-j\omega}) \begin{pmatrix} \frac{1}{2}\theta^2 & 0 \\ 0 & \frac{1}{2}\theta^2 \end{pmatrix}. \tag{1.3.12}$$

Then under the hypothesis, we have

$$\frac{\partial}{\partial\theta} \int_{-\pi}^{\pi} \Psi(\omega)^* \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \Psi(\omega)\, d\omega \Big|_{\theta=\theta_0} = \boldsymbol{0}, \tag{1.3.13}$$

which satisfies the assumption in Theorem 1.3.2.

1.3.2 Proof of Theorem 4.1

First, we derive the asymptotics of $\boldsymbol{P}_n(\boldsymbol{\theta}_0)$ and $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$.

Lemma 1.3.4. *Suppose* $\{X(t)\}_{t=0}^{\infty}$ *is generated by* (1.1.1) *satisfying* (1.2.9). *Then*

$$\boldsymbol{I}_{n,X}(\omega) = \Psi(\omega)\boldsymbol{I}_{n,Z}(\omega)\Psi(\omega)^* + \boldsymbol{R}_n(\omega). \tag{1.3.14}$$

If $\phi(\omega)$ *is a* $d \times d$ *matrix-valued continuous function on* $[-\pi,\pi]$, *then*

$$x_n \int_{-\pi}^{\pi} \mathrm{tr}[\, \boldsymbol{R}_n(\omega)\phi(\omega)\,] d\omega \xrightarrow{\mathcal{P}} 0. \tag{1.3.15}$$

Proof. We follow the proof of the univariate case in Mikosch et al. (1995).

$$\begin{aligned} d_{n,X}(\omega) &= n^{-1/\alpha}\sum_{t=1}^{n}\boldsymbol{X}(t)e^{i\omega t} = n^{-1/\alpha}\sum_{t=1}^{n}e^{i\omega t}(\sum_{j=0}^{\infty}\Psi(j)\boldsymbol{Z}(t-j)) \\ &= \Psi(\omega)d_{n,Z}(\omega) + n^{-1/\alpha}\sum_{j=0}^{\infty}\Psi(j)e^{ij\omega}\mathbf{Y}_{n,j}(\omega), \\ &= \boldsymbol{J}_{n,Z}(\omega) + n^{-1/\alpha}\mathbf{Y}_n(\omega) \quad \text{(say)}, \end{aligned}$$

where

$$\mathbf{Y}_{n,j}(\omega) = \sum_{t=1-j}^{n-j}\boldsymbol{Z}(t)e^{i\omega t} - \sum_{t=1}^{n}\boldsymbol{Z}(t)e^{i\omega t}. \tag{1.3.16}$$

Then we have

$$\boldsymbol{R}_n(\omega) = n^{-1/\alpha}\mathbf{Y}_n(\omega)\boldsymbol{J}_{n,Z}(\omega)^* + n^{-1/\alpha}\boldsymbol{J}_{n,Z}(\omega)\mathbf{Y}_n(\omega)^* + n^{-2/\alpha}\mathbf{Y}_n(\omega)\mathbf{Y}_n(\omega)^*. \tag{1.3.17}$$

$\sum_{j=0}^{\infty}\Psi(j)e^{ij\omega} \leq \sum_{j=0}^{\infty}\|\Psi(j)\| < \infty$, so that $\|\Psi(\omega)\|$ is stochastically bounded. Since every element of $\boldsymbol{Z}(t)$ is in the domain of attraction of a stable law with a parameter α, $\boldsymbol{J}_{n,Z}(\omega)$ is also stochastically bounded. As results in the proof of lemma 6.2 in Mikosch et al. (1995), we know that for each $l \in 1, 2, \dots, d$,

$$\sum_{j=0}^{\infty}\Psi(j)_{kl}e^{ij\omega}\mathbf{Y}_{n,j}(\omega)_l = O_p(1) \tag{1.3.18}$$

and

$$\int_{-\pi}^{\pi} n^{-2/\alpha}|\sum_{j=0}^{\infty}\Psi(j)_{kl}e^{ij\omega}\mathbf{Y}_{n,j}(\omega)_l|^2 d\omega = o_p(x_n^{-2}). \tag{1.3.19}$$

Combining these two results, it is easy to see that $\mathbf{Y}_n(\omega) = O_p(1)$, and by the boundedness of $\phi(\omega)$,

$$\begin{aligned}
x_n\Big|\int_{-\pi}^{\pi}\mathrm{tr}[\boldsymbol{R}_n(\omega)\phi(\omega)]\,d\omega\Big| &\leq x_n\int_{-\pi}^{\pi}|\mathrm{tr}[\,\boldsymbol{R}_n(\omega)\phi(\omega)\,]|\,d\omega \\
&\leq x_n\int_{-\pi}^{\pi}\|\boldsymbol{R}_n(\omega)\|_E\|\phi(\omega)\|_E\,d\omega \\
&\leq c_1x_n\int_{-\pi}^{\pi}\|n^{-1/\alpha}\boldsymbol{Y}_n(\omega)\boldsymbol{J}_{n,Z}(\omega)^*\|_E+\|n^{-1/\alpha}\boldsymbol{J}_{n,Z}(\omega)\boldsymbol{Y}_n(\omega)^*\|_E \\
&\qquad +\|n^{-2/\alpha}\boldsymbol{Y}_{n,Z}(\omega)\boldsymbol{Y}_{n,Z}(\omega)^*)\|_E\,d\omega \\
&\leq c_2x_n\Big\{\Big(\int_{-\pi}^{\pi}\|\boldsymbol{I}_{n,Z}(\omega)\|_E^2\,d\omega\Big)^{1/2}\Big(\int_{-\pi}^{\pi}n^{-2/\alpha}\|\boldsymbol{Y}_n(\omega)\|_E^2\,d\omega\Big)^{1/2} \\
&\qquad +\int_{-\pi}^{\pi}n^{-2/\alpha}\|\boldsymbol{Y}_n(\omega)\|_E^2\,d\omega\Big\}. \\
&\xrightarrow{\mathcal{P}} 0.
\end{aligned}$$

□

Before looking into the asymptotics of $\boldsymbol{P}_n(\boldsymbol{\theta}_0)$, we have to show the existence of the limit matrix of the autocovariance matrix in distribution. If the components of the vector $\boldsymbol{Z}$ are mutually independent, then we have the lemma due to Resnick (1986) by applying continuous mapping theorem.

Suppose $y_n = (n\log n)^{1/\alpha}$. It is obvious that $Z(1)_k$'s satisfy followings:

$$P(|Z(1)_i| > x) = x^{-\alpha}L(x), \quad i = 1, 2, \ldots, d \tag{1.3.20}$$

with $\alpha > 0$ and $L(x)$ a slowly varying function at ∞ and

$$\frac{P(Z(1)_i > x)}{P(|Z(1)_i| > x)} \to p, \quad \frac{P(Z(1)_i < -x)}{P(|Z(1)_i| > x)} \to q \tag{1.3.21}$$

as $x \to \infty$, $0 \leq p \leq 1$ and $q = 1 - p$.

Lemma 1.3.5. *Let $\{\boldsymbol{Z}(t)\}$ be a sequence of iid random vectors satisfying* (B.9) *and* (B.10) *with $0 < \alpha < 2$ and $E|Z(1)_i|^\alpha = \infty$ for all $i = 1, 2, \ldots, d$. Then*

$$\Big(n^{-2/\alpha}\sum_{t=1}^{n}\boldsymbol{Z}(t)\boldsymbol{Z}(t)', y_n^{-1}\sum_{t=1}^{n}\boldsymbol{Z}(t)\boldsymbol{Z}(t+1)', \ldots, y_n^{-1}\sum_{t=1}^{n}\boldsymbol{Z}(t)\boldsymbol{Z}(t+h)'\Big) \Rightarrow (S(0), S(1), \ldots, S(h)), \tag{1.3.22}$$

where $S(0), S(1), \ldots, S(h)$ are independent stable random matrices; the components of $S(0)$ are all positive with index $\alpha/2$, and $S(1), \ldots, S(h)$ are identically distributed with index α.

Proof of Theorem 1.3.1. From Lemma 1.3.5, we can see that

$$\left(n^{-2/\alpha}\hat{\Gamma}_{n,Z}(0), y_n^{-1}\hat{\Gamma}_{n,Z}(k), k = 1\ldots, h\right) \Rightarrow (S(0), S(1), \ldots, S(h)).$$

Note that $\operatorname{tr}\hat{\Gamma}_{n,Z}(0) = \|Z\|_N^2$, according to the continuous mapping theorem, the statement holds true if we show

$$x_n \int_{-\pi}^{\pi} \operatorname{tr}\Big[\,\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^*\}\phi_k(\omega)\Big]\, d\omega$$

$$\xrightarrow{\mathcal{L}} \sum_{i,j=1}^{d}\sum_{h=1}^{\infty} S(h)_{ij}\int_{-\pi}^{\pi}(A(\omega)+\overline{A(\omega)})_{ij}\, d\omega. \quad (1.3.23)$$

From Lemma 1.3.4 and Lemma 1.4.1 in Appendix,

$$\begin{aligned}
& x_n \int_{-\pi}^{\pi} \operatorname{tr}\Big[\{\boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^*\}\phi_k(\omega)\Big]\, d\omega \\
= \;& x_n \int_{-\pi}^{\pi} \operatorname{tr}\Big[\,\{\Psi(\omega)\boldsymbol{I}_{n,Z}(\omega)\Psi(\omega)^* - \Psi(\omega)\hat{\Gamma}_{n,Z}(0)\Psi(\omega)^* + \boldsymbol{R}(\omega)\}\phi_k(\omega)\Big]\, d\omega \\
= \;& x_n \int_{-\pi}^{\pi} \operatorname{tr}\Big[\,\{\Psi(\omega)(\boldsymbol{I}_{n,Z}(\omega) - \hat{\Gamma}_{n,Z}(0))\Psi(\omega)^*\}\phi_k(\omega)\Big]\, d\omega + x_n\int_{-\pi}^{\pi}\operatorname{tr}\left[\,\boldsymbol{R}(\omega)\phi_k(\omega)\,\right] d\omega \\
= \;& x_n \int_{-\pi}^{\pi} \operatorname{tr}\left[\left(\sum_{h=1}^{n-1}\hat{\Gamma}_{n,Z}(h)e^{-ih\omega}\right)\Psi(\omega)^*\phi_k(\omega)\Psi(\omega)\right] d\omega \\
& + x_n \int_{-\pi}^{\pi} \operatorname{tr}\left[\left(\sum_{h=1}^{n-1}\hat{\Gamma}_{n,Z}(h)'e^{ih\omega}\right)\Psi(\omega)^*\phi_k(\omega)\Psi(\omega)\right] d\omega + o_p(1) \\
= \;& x_n\left(\int_{-\pi}^{\pi}\sum_{i,j=1}^{d}\left[\sum_{h=1}^{n-1}\hat{\Gamma}_{n,Z}(h)\right]_{ij}\left[e^{-ih\omega}\Psi(-\omega)^*\phi_k(-\omega)\Psi(-\omega)\right]_{ij} d\omega\right. \\
& \left. + \int_{-\pi}^{\pi}\sum_{i,j=1}^{d}\left[\sum_{h=1}^{n-1}\hat{\Gamma}_{n,Z}(h)\right]_{ij}\left[\,e^{ih\omega}\Psi(\omega)^*\phi_k(\omega)\Psi(\omega)\,\right]_{ij} d\omega\right) \\
\xrightarrow{\mathcal{L}} \;& \sum_{i,j=1}^{d}\sum_{h=1}^{\infty} S(h)_{ij}\int_{-\pi}^{\pi}(A(\omega)+\overline{A(\omega)})_{ij}\, d\omega.
\end{aligned}$$

□

Remark 1.3.6. The last result of convergence is due to Lemma 1.4.1, which guarantees the tightness of the convergency.

Remark 1.3.7. The assumption of independence on the components of $\boldsymbol{Z}(t)$ is for simplicity and for simulation. The condition of regular variation on the vector case is crucial for the convergence of $\boldsymbol{Z}(t)$ with some other technical conditions. For detail, we recommend to refer to Bartkiewicz et al. (2010).

From the definition, we have

$$\sum_{t=1}^{n}\sum_{i=1}^{d}\tilde{Z}(t)_i^2 = 1 \quad \text{almost surely,} \tag{1.3.24}$$

which shows the second moment of $\tilde{\boldsymbol{Z}}(t)$ is finite. By the properties that the components of vectors are mutually independent and they are symmetry around 0, we assume generally

$$E\left[\tilde{Z}(t)_i\tilde{Z}(s)_i\right] = \Sigma_{\tilde{Z}} = \begin{cases} \frac{\sigma_{ij}}{n}, & \text{if } t = s, \\ 0 & \text{if } t \neq s. \end{cases} \tag{1.3.25}$$

This is not a special case since we have the following example:

Example 1 (Case that the correlation between all elements of $\boldsymbol{Z}(t)$ is 1)**.** Assume that the marginal distributions $Z(t)_j$ of $\boldsymbol{Z}(t)$ are independent symmetric α-stable distributions with different scales σ_j. Then since the sum of all marginal distribution $\sum_{j=1}^{n} Z(t)_j$ has the same distribution, we can see that

$$E\sum_{i=1}^{d}\tilde{Z}(t)_i^2 = \frac{1}{n}. \tag{1.3.26}$$

Also, according to the different scale, we can write $Z(t)_j = \sigma_j Z'(t)_j$ where all $Z'(t)_j$ are stable with scale 1. Then we have

$$E\sum_{i=1}^{d}\tilde{Z}(t)_i^2 = E\sum_{i=1}^{d}\sigma_i^2\tilde{Z}'(t)_i = \frac{1}{n}, \tag{1.3.27}$$

which is followed by

$$E\,\tilde{Z}'(t)_i = \frac{1}{n}\left(\sum_{i=1}^{d}\sigma_i^2\right)^{-1}. \tag{1.3.28}$$

Accordingly, we have

$$E(\tilde{Z}(t)_i\tilde{Z}(t)_j) = \frac{1}{n}\frac{\sigma_i\sigma_j}{\sum_{i=1}^{d}\sigma_i^2}. \tag{1.3.29}$$

The representation (1.3.23) is just a generation of this idea.

Lemma 1.3.8. *Assume the covariance matrix of self-normalized process* $\{\tilde{Z}\}$ *is given by* $\Sigma_{\tilde{Z}}$. *If* $\alpha \in [1, 2)$, *then*

$$(n^{-2/\alpha}\|Z\|_N^2)^{-2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{P}} \boldsymbol{W}, \tag{1.3.30}$$

where the (a, b)*-component of* $\boldsymbol{W}$ *satisfies*

$$W_{ab} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\text{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\right.$$
$$\left.+\text{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\text{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\right)d\omega, \tag{1.3.31}$$

where $\tilde{\boldsymbol{g}}(\omega)$ *is defined as*

$$\tilde{\boldsymbol{g}}(\omega) = \Psi(\omega)\Sigma_{\tilde{Z}}\Psi(\omega)^*. \tag{1.3.32}$$

Proof. Apply the decomposition in Lemma 1.3.4 again, we have

$$\boldsymbol{I}_{n,X}(\omega) = \Psi(\omega)\boldsymbol{I}_{n,Z}(\omega)\Psi(\omega)^* + \boldsymbol{R}_n(\omega). \tag{1.3.33}$$

Using self-normalized form, we can see that

$$\boldsymbol{I}_{n,\tilde{X}}(\omega) \equiv (n^{-2/\alpha}\|Z\|_N^2)^{-2}\boldsymbol{I}_{n,X}(\omega) = \Psi(\omega)\boldsymbol{I}_{n,\tilde{Z}}(\omega)\Psi(\omega)^* + \boldsymbol{R}_n(\omega). \tag{1.3.34}$$

Taking the expectation of the product of periodogram of $\tilde{\boldsymbol{Z}}$, we obtain

$$\begin{aligned} &E(I_{n,\tilde{Z}}(\lambda_1)_{pq}I_{n,\tilde{Z}}(\lambda_2)_{rs}) \\ &\quad= E\left(\sum_{m,l,k,j}\tilde{Z}_p(m)\tilde{Z}_q(l)\tilde{Z}_r(k)\tilde{Z}_s(j)\exp\{\,i((j-k)\lambda_1-(l-m)\lambda_2)t\}\right) \\ &\quad= \begin{cases}\sigma_{pq}\sigma_{rs}+\sigma_{pr}\sigma_{qs}+o_p(1) & \text{if } \lambda_1=\lambda_2,\\ \sigma_{pq}\sigma_{rs}+\sigma_{ps}\sigma_{qr}+o_p(1) & \text{if } \lambda_1=-\lambda_2.\end{cases} \end{aligned} \tag{1.3.35}$$

Therefore, if we write $g(\lambda)_{ab} = (\sum_{j=0}^n \Psi(j)e^{-ij\lambda})_{ab}$, then

$$\begin{aligned}\lim_{n\to\infty}E(I_{n,\tilde{X}}(\lambda_t)_{pq}I_{n,\tilde{X}}(\lambda_t)_{rs}) &= \sum_{k,l,m,n} g(\lambda_t)_{pk}\overline{g}_{ql}(\lambda_t)\overline{g}_{rm}(\lambda_t)g_{sn}(\lambda_t)(\sigma_{pq}\sigma_{rs}+\sigma_{pr}\sigma_{qs}) \\ &= \tilde{\boldsymbol{g}}(\omega)_{pq}\tilde{\boldsymbol{g}}(\omega)_{rs}+\tilde{\boldsymbol{g}}(\omega)_{pr}\tilde{\boldsymbol{g}}(\omega)_{qs}.\end{aligned}$$

If $\alpha \in [1,2)$, we can write $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ in the integral form, i.e.

$$
\begin{aligned}
&E[\,\boldsymbol{S}_n(\boldsymbol{\theta}_0)_{ab}\,] \\
&= \frac{1}{2\pi}\Big\{\int_{-\pi}^{\pi} \sum_{\beta_1,\beta_2,\beta_3,\beta_4=1}^{d} \tilde{\boldsymbol{g}}(\omega)_{\beta_1\beta_2}\tilde{\boldsymbol{g}}(\omega)_{\beta_3\beta_4} \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_2\beta_1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_4\beta_3}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} d\omega \\
&\quad + \int_{-\pi}^{\pi} \sum_{\beta_1,\beta_2,\beta_3,\beta_4=1}^{d} \tilde{\boldsymbol{g}}(\omega)_{\beta_1\beta_3}\tilde{\boldsymbol{g}}(\omega)_{\beta_2\beta_4} \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_2\beta_1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{\beta_4\beta_3}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} d\omega\Big\}.
\end{aligned}
\tag{1.3.36}
$$

In other words,

$$
\begin{aligned}
W_{ab} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Bigg(&\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right] \\
&+\mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right] \mathrm{tr}\left[\tilde{\boldsymbol{g}}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\Bigg) d\omega.
\end{aligned}
\tag{1.3.37}
$$

The convergence in probability is guaranteed by the result that

$$
\sum_{k\neq l} \mathrm{Cov}(I_{n,\tilde{Z}}(\lambda_k)^2_{pq}, I_{n,\tilde{Z}}(\lambda_l)^2_{rs}) = O(n). \tag{1.3.38}
$$

□

Corollary 1.3.9. If all elements of $\boldsymbol{Z}(t)$ are i.i.d symmetric α stable random variables, then

$$
(n^{-2/\alpha}\|Z\|_N^2)^{-2}\boldsymbol{S}_n(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{P}} \boldsymbol{W}, \tag{1.3.39}
$$

where the (a,b)-component of $\boldsymbol{W}$ is

$$
\begin{aligned}
W_{ab} = \frac{1}{2\pi d^2}\int_{-\pi}^{\pi} \Bigg(&\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right] \\
&+\mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_a}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right] \mathrm{tr}\left[\boldsymbol{g}(\omega)\frac{\partial \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}}{\partial\theta_b}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]\Bigg) d\omega.
\end{aligned}
\tag{1.3.40}
$$

Proof of Theorem 1.3.2. First, we will derive the asymptotic distribution of the empirical likelihood ratio. For convenience, we set $\boldsymbol{p} = (p_1, \ldots, p_n)$. Introducing Lagrange multiplier $L(\boldsymbol{p}, \boldsymbol{\phi}, k)$,

$$L(\boldsymbol{p}, \boldsymbol{\phi}, k) = \sum_{t=1}^{n} \log(np_t) - n\boldsymbol{\phi}' \sum_{t=1}^{n} p_t \boldsymbol{m}(\lambda_t; \boldsymbol{\theta_0}) + k\left(\sum_{t=1}^{n} p_t - 1\right). \tag{1.3.41}$$

Differentiating $L(\boldsymbol{p}, \boldsymbol{\phi}, k)$ with respect to all parameters, we have equations:

$$p_t = \frac{1}{n}\frac{1}{1+\boldsymbol{\phi}'\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0)}, \tag{1.3.42}$$

where $\boldsymbol{\phi}$ satisfies

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0)}{1+\phi'\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0)} = \boldsymbol{0}. \tag{1.3.43}$$

If we write

$$Y_t = \boldsymbol{\phi}'\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0), \tag{1.3.44}$$

then we have

$$np_t = (1+Y_t)^{-1}, \tag{1.3.45}$$

and from (1.3.43)

$$\boldsymbol{\phi} = \boldsymbol{S}_n(\boldsymbol{\theta}_0)^{-1}\left\{\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0)\right\} + \boldsymbol{\epsilon}, \tag{1.3.46}$$

where

$$\boldsymbol{\epsilon} = \frac{1}{n}\sum_{t=1}^{n} \frac{\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0)Y_t^2}{1+Y_t}, \tag{1.3.47}$$

since

$$\frac{1}{1+Y_t} = 1 - Y_t + \frac{Y_t^2}{1+Y_t}. \tag{1.3.48}$$

Thus the empirical likelihood ratio can be decomposed like

$$\begin{aligned} -2\log R(\boldsymbol{\theta}_0) &= 2\sum_{t=1}^{n} \log(1+Y_t) \\ &= 2\sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} Y_t^2 + \sum_{t=1}^{n} O_p(Y_t^3) \\ &= n\boldsymbol{P}_n(\boldsymbol{\theta}_0)'\boldsymbol{S}_n(\boldsymbol{\theta_0})^{-1}\boldsymbol{P}_n(\boldsymbol{\theta}_0) - n\boldsymbol{\epsilon}'\boldsymbol{S}_n(\boldsymbol{\theta}_0)\boldsymbol{\epsilon} + \sum_{t=1}^{n} O_p(Y_t^3). \end{aligned} \tag{1.3.49}$$

To guarantee the order of convergence, we must control the order of $\boldsymbol{\epsilon}$ and Y_t^3 simultaneously. Fortunately, we can keep them converge to 0 in probability in the stable case as well as in the usual regular case.

To give an clear overview, we define an order in the probability order notation. If $O_p(A_n)/O_p(B_n) \to o_p(1)$, then we say A_n convergences to 0 in probability faster than B_n. It is denoted by

$$\min(O_p(A_n), O_p(B_n)) = O_p(A_n) \quad \text{or} \quad O_p(A_n) \leq O_p(B_n). \tag{1.3.50}$$

If $O_p(A_n)/O_p(B_n) \to O_p(1)$, then we say that A_n is equivalent to B_n in the order sense.

Define $Z_n = \max_{1\leq k\leq n} \boldsymbol{m}(\lambda_t; \boldsymbol{\theta}_0)$. Then it is easy to see that

$$Z_n \leq \max_{1\leq k\leq n} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr}\{\boldsymbol{f}(\lambda_t; \boldsymbol{\theta}_0)^{-1} \boldsymbol{I}_n(\lambda_t)\} \right\| \leq C \sup_{\substack{i,j \\ \omega\in[-\pi,\pi]}} |\boldsymbol{I}_n(\omega)_{ij}|. \tag{1.3.51}$$

Then from (1.3.46), we have

$$O_p(\|\boldsymbol{\phi}\|)(O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}_0)\|) - O_p(Z_n)O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|)) \leq \|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|. \tag{1.3.52}$$

From (1.3.47), this leads us to the order of $\boldsymbol{\epsilon}$,

$$\begin{aligned} \|\boldsymbol{\epsilon}\| &\leq \|Z_n\| \|\boldsymbol{S}_n(\boldsymbol{\theta}_0)\| \|\boldsymbol{\phi}\|^2 |1+Y_t|^{-1} = O_p(\|Z_n\|)O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}_0)\|)O_p(\|\boldsymbol{\phi}\|^2) \\ &= \min(O_p(Z_n)O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}_0)^{-1}\|)O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|^2), O_p(Z_n^{-1})O_p(\boldsymbol{S}_n(\boldsymbol{\theta}_0))) \end{aligned} \tag{1.3.53}$$

In the regular case, we can see that

$$O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}_0)\|) = O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}_0)\|^{-1}) = O_p(1). \tag{1.3.54}$$

Then

$$\begin{aligned} &\min(O_p(Z_n)O_p(\|\boldsymbol{S}_n(\boldsymbol{\theta}_0)^{-1}\|)O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|^2), O_p(Z_n^{-1})O_p(\boldsymbol{S}_n(\boldsymbol{\theta}_0))) \\ &\qquad = O_p(Z_n)O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|^2) \end{aligned} \tag{1.3.55}$$

from LIL and CLT. Thus the order of $\|\boldsymbol{\epsilon}\|$ is

$$\|\boldsymbol{\epsilon}\| = O_p(n^{-1}\log n) \tag{1.3.56}$$

since

$$O_p(Z_n) = O_p(\log n) \tag{1.3.57}$$
$$O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|) = O_p(n^{-1/2}). \tag{1.3.58}$$

In the stable case, defining the statistic well seems a little crucial. In this paper, the order of the statistic $\boldsymbol{S}_n(\boldsymbol{\theta}_0)$ is $O_p(1)$. As shown in Mikosch and Samorodnitsky [2000], if $\alpha \neq 1$,

$$\begin{aligned} O_p(Z_n) &= O_p((\log n)^{2-2/\alpha}) && (1.3.59) \\ O_p(\|\boldsymbol{P}_n(\boldsymbol{\theta}_0)\|) &= O_p((\log n/n)^{1/\alpha}), && (1.3.60) \end{aligned}$$

which is followed by

$$\|\boldsymbol{\phi}\| = O_p((\log n/n)^{1/\alpha}). \quad (1.3.61)$$

Therefore

$$\|\boldsymbol{\epsilon}\| = O_p((\log n)^2 n^{-2/\alpha}). \quad (1.3.62)$$

In both cases, we can simplify the notation, that is,

$$\|\boldsymbol{\epsilon}\| = o_p(1). \quad (1.3.63)$$

Last, we investigate the third term in (1.3.49). From (1.3.44), it is easy to see that

$$|Y_t|^3 \leq \|\phi\|^3 \|\boldsymbol{m}(\lambda_t;\boldsymbol{\theta}_0)\|^3 = O_p((\log n)^{2+1/\alpha} n^{-3/\alpha}) \quad (1.3.64)$$

Now, multiplying true order x_n^2/n to the empirical likelihood ratio in (1.3.49), the orders of the last two terms in the right hand side are $O_p((\log n)^{4-2/\alpha} n^{-2/\alpha})$ and $O_p((\log)^{2-1/\alpha} n{-}1/\alpha)$ respectively, and thus $o_p(1)$.

Apply Theorem 1.3.1 to $x_n \boldsymbol{P}_n(\boldsymbol{\theta}_0)$, we can see that

$$\begin{aligned} x_n \boldsymbol{P}_n(\boldsymbol{\theta}_0) &= \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr} \left[\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \boldsymbol{I}_{n,X}(\omega) \right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \frac{x_n}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{tr} \left[\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \{ \boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^* \} \right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \frac{x_n}{2\pi} \begin{pmatrix} \int_{-\pi}^{\pi} \mathrm{tr} \left[\frac{\partial}{\partial \theta_1} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \{ \boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^* \} \right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \int_{-\pi}^{\pi} \mathrm{tr} \left[\frac{\partial}{\partial \theta_2} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \{ \boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^* \} \right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ \vdots \\ \int_{-\pi}^{\pi} \mathrm{tr} \left[\frac{\partial}{\partial \theta_q} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \{ \boldsymbol{I}_{n,X}(\omega) - \Psi(\omega)(\hat{\Gamma}_{n,Z}(0))\Psi(\omega)^* \} \right] d\omega \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \end{pmatrix} \\ &\xrightarrow{\mathcal{L}} \frac{1}{2\pi} \sum_{i,j=1}^{d} \sum_{h=1}^{\infty} S(h)_{ij} \begin{pmatrix} \int_{-\pi}^{\pi} (B_1(\omega) + \overline{B_1(\omega)})_{ij} d\omega \\ \int_{-\pi}^{\pi} (B_2(\omega) + \overline{B_2(\omega)})_{ij} d\omega \\ \vdots \\ \int_{-\pi}^{\pi} (B_q(\omega) + \overline{B_q(\omega)})_{ij} d\omega. \end{pmatrix} \end{aligned}$$

where

$$B_k(\omega) = \Psi(\omega)^* \frac{\partial}{\partial \theta_k} \boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1} \Psi(\omega) \quad k = 1, \ldots, q.$$

Remember that

$$n^{-2/\alpha}\|Z\|_N^2 \xrightarrow{\mathcal{L}} S_{\alpha/2}, \tag{1.3.65}$$

$$\frac{x_n \boldsymbol{P}_n(\boldsymbol{\theta}_0)}{n^{-2/\alpha}\|Z\|_N^2} \xrightarrow{\mathcal{L}} \frac{1}{2\pi}\sum_{i,j=1}^{d}\sum_{h=1}^{\infty}\frac{S(h)_{ij}}{S_{\alpha/2}}\begin{pmatrix}\int_{-\pi}^{\pi}(B_1(\omega)+\overline{B_1(\omega)})_{ij}d\omega\\ \int_{-\pi}^{\pi}(B_2(\omega)+\overline{B_2(\omega)})_{ij}d\omega\\ \dots\\ \int_{-\pi}^{\pi}(B_q(\omega)+\overline{B_q(\omega)})_{ij}d\omega\end{pmatrix}. \tag{1.3.66}$$

Thus the limit of $-2\frac{x_n^2}{n}\log R(\theta_0)$ is

$$-2\frac{x_n^2}{n}\log R(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \boldsymbol{V}'\boldsymbol{W}^{-1}\boldsymbol{V},$$

where

$$\boldsymbol{V} = \frac{1}{2\pi}\sum_{i,j=1}^{d}\sum_{h=1}^{\infty}\frac{S(h)_{ij}}{S_{\alpha/2}}\begin{pmatrix}\int_{-\pi}^{\pi}(B_1(\omega)+\overline{B_1(\omega)})_{ij}d\omega\\ \int_{-\pi}^{\pi}(B_2(\omega)+\overline{B_2(\omega)})_{ij}d\omega\\ \dots\\ \int_{-\pi}^{\pi}(B_q(\omega)+\overline{B_q(\omega)})_{ij}d\omega,\end{pmatrix} \tag{1.3.67}$$

and the (a, b)-element of $\boldsymbol{W}$ is represented in the Theorem 1.3.1. □

1.4 Appendix

1.4.1 Tightness

Lemma 1.4.1 (Klüppelberg and Mikosch (1996)). *Suppose $\{Z_t\}_{1\le t\le n}$ is a sequence of iid symmetric α-stable random variables for $\alpha \in (0,2)$. Let f_t be real numbers such that*

$$\sum_{t=-\infty}^{\infty}|f_t|^{\mu} < \infty \tag{1.4.1}$$

for some $\mu < \alpha$. If $f_0 = 0$, then

$$(\gamma_{n,Z}^2, y_n^{-1}\sum_{1\le t,s\le n} f_{t-s}Z_tZ_s) \to_d (Y_0, Z_1(\sum_{t=1}^{\infty}|f_t + f_{-t}|^{\alpha})^{1/\alpha}). \tag{1.4.2}$$

If $f_0 \neq 0$, then

$$n^{-2/\alpha}\sum_{1\le t,s\le n} f_{t-s}Z_tZ_s \to_d f_0Y_0. \tag{1.4.3}$$

1.4.2 Regularly Varying Tail

Definition 1.4.2. A distribution F has exponential tails with rate $\alpha > 0$, if

$$\lim_{t\to\infty} \frac{\bar{F}(t-u)}{\bar{F}(t)} = e^{\alpha u} \quad \text{for all real } u. \tag{1.4.4}$$

It is denoted by $F \in \mathcal{L}_\alpha$.

The definition is equivalent to the definition of regularly varying tail, if one put $\log t$ and $\log u$ into the definition above. To guarantee $\|Z\|_N^2$ is in the domain of attraction of a stable limit with $\alpha/2$, we have the Theorem below.

Theorem 1.4.3 (Embrechts and Goldie (1980, Theorem 3))**.** *If both $F \in \mathcal{L}_\alpha$, $G \in \mathcal{L}_\alpha$, then $H = F * G \in \mathcal{L}_\alpha$.*

Remark 1.4.4. Regularly varying tail is necessary and sufficient condition for a sequence of i.i.d random variables or random vectors in the domain of attraction of a stable law.

1.4.3 Estimation of autocorrelation

To estimate $\Gamma(0)^{-1}\Gamma(j)$, we can see the problem as a fitting problem, i.e. fit the spectral whose inverse is

$$(I - \Theta e^{-ij\omega})(I - \Theta' e^{ij\omega}). \tag{1.4.5}$$

Then the pivotal value satisfies

$$\frac{\partial}{\partial \Theta} \int_{-\pi}^{\pi} (I - \Theta e^{-ij\omega})(I - \Theta' e^{ij\omega}) \boldsymbol{g}(\omega)\, d\omega|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}. \tag{1.4.6}$$

In fact, using formula

$$\frac{\partial}{\partial \Theta} \mathrm{tr}(\Theta \boldsymbol{g}(\omega)) = \Theta' \boldsymbol{g}(\omega) \tag{1.4.7}$$

$$\frac{\partial}{\partial \Theta} \mathrm{tr}(\Theta\Theta' \boldsymbol{g}(\omega)) = \boldsymbol{g}(\omega)\Theta + \boldsymbol{g}(\omega)\Theta', \tag{1.4.8}$$

we have

$$\Gamma(0)\,\Theta_0 = \Gamma(j) \tag{1.4.9}$$

by Herglotz's spectral representation theorem. That is

$$\Gamma(j) = \int_{-\pi}^{\pi} e^{ij\omega}\, dF(\omega). \tag{1.4.10}$$

Thus we have

$$\Theta_0 = \Gamma(0)^{-1}\Gamma(j). \tag{1.4.11}$$

Also, we can extend this full autocorrelation case to any one element case in the same way. (See Numerical Study above.)

1.4.4 Program

As an example, the program with the true $\theta_0 = 0$ is given below. " $-$ " shows the code depending on the program.

```
Clear["Global`*"]
Combine[P__] := MapThread[List, {P}];
varfil[P__] := Reverse[{P}];
order[P__] := Length[varfil[P]];
innov[s_, P__] := Join[ConstantArray[0, {order[P] - 1, 2}], s];
VAR[s_, P__] := (Clear[X]; X = {s[[1]]}; X = innov[X, P];
  For[i = 1, i < Length[s], i++,
   X = Append[X,
     Total[MapThread[Dot, {varfil[P], Take[X, {i, i + order[P] - 1}]}]] +
      s[[i + 1]]]]; Drop[X, order[P] - 1])

T = 300;
z1 = RandomVariate[StableDistribution[1, 1.5, 0, 0, 1], T];
z2 = RandomVariate[StableDistribution[1, 1.5, 0, 0, 1], T];
z = Combine[z1, z2];
A = {{0.7, 0}, {0.1, 0.5}};
Eigenvalues[A]
VAR[z, A];
d = Length[X[[1]]];
l = 1;
e[r_, n_] := Table[E^(2 Pi I (r - 1) (s - 1)/T), {s, 1, n}]
Periodogram[x_, r_, a_] :=
 T^(-2/a) Outer[Times, e[r, T].N[x], Conjugate[e[r, T].N[x]]]
filter[P__] := -{-IdentityMatrix[d], P}
fitar[P__, r_] := (F = filter[P];
  Transpose[Total[MapThread[Times, {F, Conjugate[e[r, Length[F]]]}]]].Total[
    MapThread[Times, {F, e[r, Length[F]]}]])
M[x_, P_, r_, a_] := D[Tr[fitar[P, r].Periodogram[x, r, a]], t1]
MList[x_, P_, tr_, a_] := Chop[Table[M[x, P, r, a], {r, 1, T}] /. {t1 -> tr}]
B = {{0.7, t1}, {0.1, 0.5}}
ec[w_, n_] := Table[E^(I (s - 1) w), {s, 1, n}]
Spectral[x_, P__] := (G = filter[P];
  Inverse[
   Transpose[Total[MapThread[Times, {G, ec[-x, Length[G]]}]]].Total[
     MapThread[Times, {G, ec[x, Length[G]]}]]])
InvSpectral[x_, P__] := (G = filter[P];
  Transpose[Total[MapThread[Times, {G, ec[-x, Length[G]]}]]].Total[
```

```
    MapThread[Times, {G, ec[x, Length[G]]}]])

tr = Chop[
  t1 /. (FindRoot[
      D[Integrate[Tr[InvSpectral[x, B].Spectral[x, A]], {x, -Pi, Pi}], t1] ==
       0, {t1, 0.701}])[[1]]]

Mtrue = MList[X, B, tr, 1.5];
p = Total[Mtrue]/T
q = Total[Mtrue^2]/T
R = (T/Log[T])^(2/1.5) (p^2/q)
w13 = Simplify[(D[InvSpectral[x, B], t1] /. {t1 -> tr})].Spectral[x, A];
w = Integrate[Tr[w13.w13] + Tr[w13] Tr[w13], {x, -Pi, Pi}]/(2 Pi)
v := Tr[w13]
VB[a_] := (k = 1; S = -1; s = 1;
  While[(s > S/1000 || s == 0) && k < 20, S = S + s;
   s = Abs[Integrate[v Cos[k x], {x, -Pi, Pi}]]^(a); Print[s]; k++]; S)
V[a_] := (VB[a])^(1/a)/Pi
ko[a_] := If[a == 1, 2/Pi, (1 - a)/(Gamma[2 - a] Cos[Pi a/2])];
Y0[a_] := StableDistribution[a/2, 1, 0, (ko[a/2])^(-2/a)];
Z0[a_] := StableDistribution[a, 0, 0, (ko[a])^(-1/a)];
Quant = Quantile[(RandomVariate[Z0[1.5], 10000]/
     RandomVariate[Y0[1.5], 10000])^2, 0.90]
Val = V[1.5]
confi = Quant Val^2/w
{0, confi}
g1 = Plot[{(T/Log[T])^(2/
       1.5) ((Total[MList[X, B, t, 1.5]]/T)^2/(Total[MList[X, B, t, 1.5]^2]/
        T)), confi}, {t, -1, 1}]
FindRoot[(T/Log[T])^(2/
      1.5) ((Total[MList[X, B, t, 1.5]]/T)^2/(Total[MList[X, B, t, 1.5]^2]/
       T)) == confi, {t, 0.1}]
FindRoot[(T/Log[T])^(2/
      1.5) ((Total[MList[X, B, t, 1.5]]/T)^2/(Total[MList[X, B, t, 1.5]^2]/
       T)) == confi, {t, -0.2}]
Show[g1, PlotRange -> {{-0.1, 0.1}, {0, 1}}]

a = Chop[t /. %---[[1]]]
b = Chop[t /. %---[[1]]]
b - a
```

Chapter 2

Asymptotic Moments of the Self-normalized Sum

Abstract

We give a general and explicit formula for the moments of the limiting distribution of symmetric self-nomalized sum of i.i.d random variables, which belong to the domain of attraction of a stable law. The result shows that the finite order moments for symmetric self-normalized sums are always finite. As an application, tail index can be estimated through our result by using moment estimators.

2.1 Introduction and preliminaries

The self-normalized method has been focused on in these two decades, and many interesting results are obtained. (See Logan et al. [1973], Griffin and Mason [1991], Klüppelberg and Mikosch [1996], Peña et al. [2009].) In this paper, we extend the result for the moments of symmetric self-normalized sum in Logan et al. [1973] to a more explicit one.

Consider a sequence $\{X_i\}_{i=1,\ldots,n}$ that X_i's are assumed to be independent and identically distributed and belong to the domain of attraction of a stable law G, the parameter of attracting stable law G is denoted by α. More specifically, we assume that the density function g of the stable distribution G satisfies

$$x^{\alpha+1}g(x) \to r, \quad x^{\alpha+1}g(-x) \to l, \tag{2.1.1}$$

where $0 < \alpha < 2$, $r + l > 0$. Also U_n and V_n^p are defined as

$$U_n = \frac{X_1 + \cdots + X_n}{n^{1/\alpha}} \tag{2.1.2}$$

and

$$V_n^2 = \frac{|X_1|^2 + \cdots + |X_n|^2}{n^{2/\alpha}}. \tag{2.1.3}$$

To have the limiting distribution of $S_n(2)$ $(= U_n/V_n)$ exist, we further assume that

$$EX_i = 0 \quad \text{if } 1 < \alpha < 2. \tag{2.1.4}$$

The limiting distribution of $S_n(2)$ is denoted by $S(2)$.

It is shown in Logan et al. [1973] that if $\alpha \neq 1$, the moments of $S(2)$ can be derived from

$$\frac{1}{\pi}\int_0^\infty \varphi(t)e^{-st}\,dt = \int_0^\infty e^{-s^2t^2/2}\mathcal{D}(t)\,dt, \tag{2.1.5}$$

where

$$\varphi(t) = Ee^{iS(2)t} = \lim_{n\to\infty} Ee^{iS_n(2)t}, \tag{2.1.6}$$

the characteristic function of the limiting distribution of $S(2)$, and

$$\mathcal{D}(t) = (1-\alpha)(2\pi^{-3})^{1/2}\frac{rD_{\alpha-2}(-it) + lD_{\alpha-2}(it)}{rD_\alpha(-it) + lD_\alpha(it)}, \tag{2.1.7}$$

$D_\nu(z)$ $(z \in \mathbb{C})$ is the parabolic cylinder functions. (See Magnus and Oberhettinger [1954].) Here are two important properties of parabolic cylinder functions for calculation.

$$\frac{d}{dz}D_\nu(z) - \frac{z}{2}D_\nu(z) + D_{\nu+1}(z) = 0; \tag{2.1.8}$$

$$\frac{d}{dz}D_\nu(z) + \frac{z}{2}D_\nu(z) - \nu D_{\nu-1}(z) = 0. \tag{2.1.9}$$

The calculation of the moments depends on the expansion of (2.1.5). We first decompose the left hand side into the form of power series.

$$\begin{aligned}
\frac{1}{\pi}\int_0^\infty \varphi(t)e^{-st}\,dt &= \frac{1}{\pi}\int_0^\infty \sum_{k=0}^\infty \frac{\varphi^{(k)}(0)}{k!}t^k e^{-st}\,dt \\
&= \sum_{k=0}^\infty \frac{\varphi^{(k)}(0)}{k!}\frac{1}{\pi}\int_0^\infty t^k e^{-st}\,dt \\
&= \sum_{k=0}^\infty \frac{1}{\pi}\varphi^{(k)}(0)s^{-k-1}. \qquad (2.1.10)
\end{aligned}$$

Secondly, the right hand side of (2.1.5) can be written as follows.

$$\begin{aligned}\int_0^\infty e^{-s^2t^2/2}\mathcal{D}(t)dt &= \sum_{k=0}^\infty \frac{\mathcal{D}^{(k)}(0)}{k!}\int_0^\infty e^{-s^2t^2/2}t^k\,dt \\ &= \sum_{k=0}^\infty \frac{\mathcal{D}^{(k)}(0)}{k!}\int_0^\infty e^{-u}\left(\frac{2u}{s^2}\right)^{\frac{k}{2}}\cdot\frac{1}{s\sqrt{2u}}\,du \\ &= \sum_{k=0}^\infty \frac{\mathcal{D}^{(k)}(0)}{k!}2^{\frac{k-1}{2}}\Gamma(\frac{k+1}{2})s^{-k-1}. \qquad (2.1.11)\end{aligned}$$

Equating coefficients of like powers of s^{-1} in (2.1.10) and (2.1.11), we can see that

$$E(S(2)^k) = i^k\varphi^{(k)}(0) = \frac{\mathcal{D}^{(k)}(0)}{k!}2^{\frac{k-1}{2}}\Gamma(\frac{k+1}{2})\pi. \qquad (2.1.12)$$

The main purpose of this paper is to derive a general and explicit formula to calculate the moments of the distribution $S(2)$ when $r = l$.

2.2 The Main result

We assume X_i's are symmetric, i.e., $r = l$, then $\mathcal{D}(t)$ becomes

$$\mathcal{D}(t) = (1-\alpha)(2\pi^{-3})^{1/2}\frac{D_{\alpha-2}(-it)+D_{\alpha-2}(it)}{D_\alpha(-it)+D_\alpha(it)} \equiv (1-\alpha)(2\pi^{-3})^{1/2}\mathcal{A}(t). \quad (2.2.1)$$

For symmetry case, we simplify the notation of the limiting distribution $S(2)$ by S.

Theorem 2.2.1. *Let S be defined above. Then for any $m = 1, 2, \ldots$, $E(S^{2m-1}) = 0$ and*

$$E(S^{2m}) = \frac{(2m-1)!!}{(2m)!}2\{(D_{\alpha-2}^{(2m)}(0)-D_\alpha^{(2m)}(0))-(1-\alpha)\sum_{k=0}^{m-1}\frac{(-1)^{m-k}\mathcal{A}^{(2k)}(0)}{(2(m-k))!!(2k)!}D_\alpha^{2(m-k)}(0)\}, \qquad (2.2.2)$$

where $\mathcal{A}^{(2k)}(0)$ satisfies $\mathcal{A}(0) = D_{\alpha-2}(0)/D_\alpha(0)$ and

$$\mathcal{A}^{(2k)}(0) = \frac{(-1)^k}{1-\alpha}2(D_{\alpha-2}^{(2k)}(0)-D_\alpha^{(2k)}(0))-2\sum_{l=0}^{k-1}\frac{\mathcal{A}^{(2l)}(0)}{(2(k-l))!!(2l)!}(-1)^{(k-l)}D_\alpha^{2(k-l)}(0)\}. \qquad (2.2.3)$$

Furthermore, suppose

$$D_\nu^{(2k)}(0) = \eta^0(k) + \sum_{j=1}^k \eta^j(k)\nu^j, \qquad (2.2.4)$$

then $\eta^j(k)$ *satisfies*

$$\begin{cases} \eta^j(k) = -\sum_{t=0}^{k-j}\binom{k-t}{j}\nu^{k-t}(k-1) + \frac{2k-1}{2}\eta^j(k-1) \quad \text{for } j\geq 0; \\ \nu^1(k) = k!\left((-1)^k + \sum_{l=1}^{k}\frac{(2l-1)!!(-1)^{k-l}}{2^l l!}\right); \\ \nu^j(k) = -\sum_{l=1}^{k-j+2}\frac{(2k-1)!!}{(2k-(2l-1))!!2^{l-1}}\nu^{j-1}(k-l) - k\nu^j(k-1) \quad \text{for } j\geq 2; \\ \nu^1(0)=1,\ \nu^j(0)=0 \ \text{ for } j\geq 2; \quad \nu^0(k)=0 \ \text{ for any } k\geq 0; \\ \eta^j(k)=0 \ \text{ for } j>k; \quad \eta^k(k)=(-1)^k \ \text{ for any } k\geq 0. \end{cases} \tag{2.2.5}$$

Corollary 2.2.2. The finite order moments for self-normalized sum of i.i.d random variables in the domain of attraction of a stable law are always finite.

2.3 Proof of Theorem 2.2.1

Set

$$A_\nu(t) = D_\nu(-it) + D_\nu(it), \tag{2.3.1}$$

then

$$\mathcal{A}(t) = \frac{D_{\alpha-2}(-it) + D_{\alpha-2}(it)}{D_\alpha(-it) + D_\alpha(it)} = \frac{A_{\alpha-2}(t)}{A_\alpha(t)}. \tag{2.3.2}$$

From (2.1.12), the moment of the limiting distribution can be simply written as

$$E(S^k) = \frac{(k-1)!!}{k!}i^k(1-\alpha)\mathcal{A}^{(k)}(0). \tag{2.3.3}$$

Note that $\frac{d}{dt}D_\nu(-it) = -i\frac{d}{dz}D_\nu(z)\Big|_{z=-it}$ and $\frac{d}{dt}D_\nu(it) = i\frac{d}{dz}D_\nu(z)\Big|_{z=it}$, it is obvious that

$$A_k(\nu) \equiv \frac{d^k}{dt^k}A_\nu(t)\Big|_{t=0} = \begin{cases} 0 & \text{if k is odd;} \\ (-1)^{k/2}2D_\nu^{(k)}(0), & \text{if k is even,} \end{cases} \tag{2.3.4}$$

where $D_\nu^{(k)}(0) = \frac{d^k}{dz^k}D_\nu(z)\Big|_{z=0}$. To prove the first statement, we use a recursive formula for nth derivative.

Lemma 2.3.1 (Xenophontos [2007])**.**

$$\left(\frac{u(x)}{v(x)}\right)^{(n)} = \frac{1}{v(x)}\left(u^{(n)}(x) - n!\sum_{j=1}^{n}\frac{v(x)^{(n+1-j)}}{(n+1-j)!(j-1)!}\left(\frac{u(x)}{v(x)}\right)^{(j-1)}\right). \tag{2.3.5}$$

Applying the formula to the case that $u(x) = A_{\alpha-2}(x)$ and $v(x) = A_\alpha(x)$, we have

$$\mathcal{A}^{(2k)}(0) = \frac{1}{1-\alpha}(A_{2k}(\alpha-2) - A_{2k}(\alpha)) - \sum_{l=0}^{k-1}\binom{2k}{2l}A_{2(k-l)}(\alpha)\mathcal{A}^{(2l)}(0). \quad (2.3.6)$$

The first result is straightforward from (2.3.3).

Next, we show the second half of Theorem 2.2.1. Differentiating (2.1.8) and (A.14) iteratively, we have

$$D_\nu^{(k)}(z) - \frac{z}{2}D_\nu^{(k-1)}(z) - \frac{k-1}{2}D_\nu^{(k-2)}(z) + D_{\nu+1}^{(k-1)}(z) = 0; \quad (2.3.7)$$

$$D_\nu^{(k)}(z) + \frac{z}{2}D_\nu^{(k-1)}(z) + \frac{k-1}{2}D_\nu^{(k-2)}(z) - \nu D_{\nu-1}^{(k-1)}(z) = 0. \quad (2.3.8)$$

Thus $D_\nu^{(k)}(0)$ can be derived from

$$D_\nu^{(k)}(0) = \frac{k-1}{2}D_\nu^{(k-2)}(0) - D_{\nu+1}^{(k-1)}(0); \quad (2.3.9)$$

$$D_\nu^{(k)}(0) = -\frac{k-1}{2}D_\nu^{(k-2)}(0) + \nu D_{\nu-1}^{(k-1)}(0). \quad (2.3.10)$$

In the case when k is odd, rewrite $2k+1$ as k, then

$$\begin{aligned} D_\nu^{(2k+1)}(0) &= -kD_\nu^{(2k-1)}(0) + \nu\left(\frac{2k-1}{2}D_{\nu-1}^{(2k-2)}(0) - D_\nu^{(2k-1)}(0)\right) \\ &= -kD_\nu^{(2k-1)}(0) - \nu\sum_{l=1}^{k}\frac{(2k-1)!!}{(2k-2l+1)!!2^{l-1}}D_\nu^{(2k-2l+1)}(0) + \nu\frac{(2k-1)!!}{2^k}D_{\nu-1}(0). \end{aligned}$$

This is a recurrence formula for $D_\nu^{(2k+1)}(0)$. If we can expand it, then it must be the product of a polynomial of ν and $D_{\nu-1}(0)$. Let $\nu^j(k)$ denote the coefficient of ν^j in the case of $(2k+1)$th derivative.

For the initial values, we can see that $\nu^1(0) = 1$, $\nu^j(0) = 0$ for all $j \geq 2$ and $\nu^0(k) = 0$ for all $k \geq 0$ from (A.14). After some painful calculation, we have

$$\nu^1(k) = k!\left((-1)^k + \sum_{l=1}^{k}\frac{(2l-1)!!(-1)^{k-l}}{2^l l!}\right); \quad (2.3.11)$$

$$\nu^j(k) = -\sum_{l=1}^{k-j+2}\frac{(2k-1)!!}{(2k-(2l-1))!!2^{l-1}}\nu^{j-1}(k-l) - k\nu^j(k-1) \quad \text{for } j \geq 2. \quad (2.3.12)$$

From the recurrence formula, one can see that the highest degree of the polynomial is $k+1$, which can be shown by the induction. Using this property reversely, one also can see that $\nu^j(k) = 0$ for any j and k satisfying $j \geq k+2$.

Corollary 2.3.2.

$$\nu^{k+1}(k) = (-1)^k, \quad \nu^k(k) = (-1)^k \frac{k}{2}. \tag{2.3.13}$$

Proof. Applying this result to (2.3.12),

$$\nu^{k+1}(k) = -\nu^k(k-1) \tag{2.3.14}$$

holds, and since the initial value $\nu^1(0) = 1$, we have

$$\nu^{k+1}(k) = (-1)^k. \tag{2.3.15}$$

Also applying the result to $\nu^k(k)$,

$$\nu^k(k) = -\nu^{k-1}(k-1) - \frac{1}{2}, \tag{2.3.16}$$

which implies

$$\nu^k(k) = (-1)^k \frac{k}{2}, \tag{2.3.17}$$

since $\nu^0(0) = 0$. □

On the other hand, when k is even, rewrite $2k+2$ as k and we have

$$D_{\nu}^{(2k+2)}(0) \quad = \quad \frac{2k+1}{2} D_{\nu}^{(2k)}(0) - D_{\nu+1}^{(2k+1)}(0)$$

Here, let $\eta^j(k)$ denote the coefficient of ν^j in the case of $2k$th derivative. Then we have

$$\eta^j(k) = -\sum_{t=0}^{k-j} \binom{k-t}{j} \nu^{k-t}(k-1) + \frac{2k-1}{2} \eta^j(k-1) \quad \text{for } j \geq 0. \tag{2.3.18}$$

From (2.3.15), $\eta^j(k) = 0$ if $j > k$ and $\eta^k(k) = -\nu^k(k-1) = (-1)^k$. □

2.4 Examples

2.4.1 Mathematica code

This section provides Mathematica code. The functions $f(j,k)$ and $g(j,k)$ denote the function $\nu^j(k)$ and $\eta^j(k)$ in the previous section, respectively. The function $A(n,a)$ is corresponding to $1/2\ A_{2n}(\alpha)$, while $CA(n,a)$ represents the function $\mathcal{A}^{(2n)}(0)$ above. Lastly, function $M(n,a)$ indicates the $2n$th moment of the limit distribution S.

```
f[1, k_] :=
  k! ((-1)^k + Sum[(2 l - 1)!! (-1)^(k - l)/(2^l l!), {l, 1, k}]);
f[j_, k_] :=
  If[j == 0, 0,
   If[k <= -1, 0,
    If[k > -1 && j >= 2,
      -(2 k - 1)!! Sum[
        f[j - 1, k - l]/((2 k - (2 l - 1))!! 2^(l - 1)), {l, 1, k - j + 2}]
          - k f[j, k - 1]]]];
g[j_, k_] :=
  If[j > k, 0,
   If[j == k, (-1)^k,
    If[k > 0, -Sum[
        Binomial[k - t, j] f[k - t, k - 1], {t, 0, k - j}] + (2 k -
         1) g[j, k - 1]/2, 0]]];
A[n_, a_] := (-1)^n g[0, n] + (-1)^n Sum[g[t, n] a^t, {t, 1, n}];
CA[n_, a_] :=
  If[n > 0, (A[n, a - 2] - A[n, a])/(1 - a) -
    Sum[CA[t, a] Binomial[2 n, 2 t] A[n - t, a], {t, 1, n - 1}], 0];
M[n_, a_] :=
  Simplify[(2 n - 1)!!/(2 n)! (-1)^n (1 - a) CA[n, a], a > 0];
```

2.4.2 Some results and knowledge

Using the code above, we obtain the general result for the moments of symmetric self-normalized moments and some special cases of $\alpha = 0.5$, $\alpha = 1.5$ and $\alpha = 2$.

Table 2.1: The $2k(k = 1, \ldots, 6)$ th moments of symmetric self-normalized sum for the case of $\alpha = 0.5, 1.5, 2$

k	$E(S^{2k})$	$\alpha = 0.5$	$\alpha = 1.5$	$\alpha = 2$
1	1	1	1	1
2	$1 + \alpha$	1.5	2.5	3
3	$1 + 3\alpha + 2\alpha^2$	3	10	15
4	$1/3(3 + 20\alpha + 34\alpha^2 + 17\alpha^3)$	7.875	55.625	105
5	$1/3(3 + 40\alpha + 130\alpha^2 + 155\alpha^3 + 62\alpha^4)$	26.25	397.5	945
6	$1/15(15 + 383\alpha + 2118\alpha^2 + 4514\alpha^3 + 4146\alpha^4 + 1382\alpha^5)$	106.838	3471.56	10395

When $\alpha = 2$, the limiting distribution S is standard normal distribution. From

Table 2.1, we can see the result is corresponding to the moments we can obtain from other methods.

Hill's estimator is proposed to be an estimator for tail index. As an alternative to it, we can apply the result above to the derivation of the tail index after calculating the asymptotic moments for the self-normalized sums by moment estimators.

Auxiliary results (1) — $A_\nu^{(k)}(0)$

$$\begin{aligned}
A_\nu(0) &= \frac{\Gamma(\frac{1}{2})2^{\frac{\nu+2}{2}}}{\Gamma(\frac{1-\nu}{2})} \\
A_\nu^{(1)}(0) &= 0 \\
A_\nu^{(2)}(0) &= \frac{2\nu+1}{2}A_\nu(0)(\equiv A_2(\nu)A_\nu(0)\,) \\
A_\nu^{(3)}(0) &= 0 \equiv A_3(\nu) \\
A_\nu^{(4)}(0) &= \frac{4\nu^2+4\nu+3}{4}A_\nu(0)(\equiv A_4(\nu)A_\nu(0)\,) \\
A_\nu^{(5)}(0) &= 0 \equiv A_5(\nu) \\
A_\nu^{(6)}(0) &= \frac{8\nu^3+12\nu^2+34\nu+15}{8}A_\nu(0)(\equiv A_6(\nu)A_\nu(0)\,).
\end{aligned}$$

Auxiliary results (2) — $\mathcal{A}^{(k)}(0)$

$$\begin{aligned}
\mathcal{A}(0) &= \frac{A_{\alpha-2}(0)}{A_\alpha(0)} = \frac{1}{1-\alpha} \\
\mathcal{A}^{(1)}(0) &= 0 \\
\mathcal{A}^{(2)}(0) &= -\frac{2}{1-\alpha} \\
\mathcal{A}^{(3)}(0) &= 0 \\
\mathcal{A}^{(4)}(0) &= \frac{8\alpha+8}{1-\alpha} \\
\mathcal{A}^{(5)}(0) &= 0 \\
\mathcal{A}^{(6)}(0) &= \frac{-96\alpha^2-144\alpha-48}{1-\alpha}.
\end{aligned}$$

Chapter 3

Rank-based Method

3.1 Introduction

3.1.1 ARMA model

The process $\{X_t; t \in \mathbb{Z}\}$ which satisfies

$$X_t - \sum_{i=1}^{p} a_i X_{t-i} = e_t + \sum_{i=1}^{q} b_i e_{t-i} \quad \text{for all } t \in \mathbb{R},$$

are concerned in the paper. Denote the density function of e_t by $f(x)$, and the common distribution of $(e_{1-q}, \dots, e_0; X_{1-p}, \dots, X_n)$ by $g_n(\cdot;\theta)$, where $\theta \in \Theta$ is the underlying parameter.

Remark 3.1.1. One can see that

$$g_n(\cdot;\theta) = g_0(e_{1-q}, \dots, e_0; X_0; \theta) \prod_{t=1}^{n} f(e_t\{e_{1-q}, \dots, X_t\}),$$

where

$$e_t\{e_{1-q}, \dots, X_t\} = \sum_{k=1}^{t} \beta_{k-1} \left(- \sum_{i=0}^{p} a_i X_{t+1-k-i} \right) + \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{t+s-k} b_k \right).$$

Remark 3.1.2. Using the representation of the common distribution above and Lemma 2.2 in Appendix, we have

$$\frac{dP_{n,\theta}}{dP_{n,\theta_0}} = \frac{g_0(e_{1-q}, \dots, X_0; \theta)}{g_0(e_{1-q}, \dots, X_0; \theta_0)} \prod_{j=1}^{n} \frac{f(e_j^0 - (\theta - \theta_0)' Z(j-1; \theta, \theta_0))}{f(e_j^0)}.$$

With the following additional abbreviation,

$$\phi_j^2(\theta_0, \theta) = \frac{f(e_j(\theta_0) - (\theta - \theta_0)' Z(j-1; \theta, \theta_0))}{f(e_j(\theta_0))},$$

we have

$$\log \frac{dP_{n,\theta}}{dP_{n,\theta_0}} = \log \frac{g_0(e_{1-q}, \ldots, X_0; \theta)}{g_0(e_{1-q,\ldots,X_0}; \theta_0)} + 2 \sum_{j=1}^{n} \log \phi_j(\theta_0, \theta).$$

3.1.2 Assumptions

Assumption for stationary and invertibility and etc.

(S1) The polynomials $A(z) = 1 + \sum_{i=1}^{p} -a_i z^i$ and $B(z) = 1 + \sum_{i=1}^{q} b_i z^q$ have no zeros with magnitude less or equal to one.

(S2) The two polynomials have no zeros in common and $a_p \neq 0$ or $b_q \neq 0$.

Assumptions for LAN

(A1) $f(x) > 0$, $x \in \mathbb{R}$; $\int_{-\infty}^{\infty} x f(x)\, dx = 0$; $\int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2 < \infty$;

(A2) f is absolutely continuous on finite intervals, i.e., there exists $\dot{f}$ such that for all $-\infty < a < b < \infty$, $f(b) - f(a) = \int_a^b \dot{f}(x)\, dx$;

(A3) letting $\phi_f \equiv -\frac{\dot{f}}{f}$, the generalized Fisher information $\int_{-\infty}^{\infty} \phi_f^2(x) f(x)\, dx \equiv \mathcal{I}_f = \sigma^{-2} \mathcal{I}_{f_1}$ is finite.

(A4) the score function ϕ_f is piecewise Lipschitz, i.e., there exist a finite partition of $\mathbb{R}$ into nonoverlapping intervals $J_1, \ldots, J_k$ and a constant A_f such that

$$|\phi_f(x) - \phi_f(y)| \leq A_f |x - y| \qquad \forall x, y \in J_i, \forall i = 1, \ldots, k.$$

3.1.3 Interpretation of the assumptions

(A1)—(A3) The LAN result holds under these assumptions;

(A4) This assumption induces that the influence of starting values on residual autocorrelations is to be asymptotically negligible. In other words, it can be shown that

$$E_\theta |\, \Delta_n(\theta) - \hat{\Delta}_n(\theta)\, | = o(1)$$

holds true, where

$$\hat{\Delta}_n(\theta) := \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(\hat{e}_j(\theta)) \sum_{k=1}^{j} \beta_{k-1} \begin{pmatrix} Y(j-k) \\ \hat{E}(j-k;\theta) \end{pmatrix},$$

and

$$\hat{e}_t := \sum_{k=1}^{t} \beta_{k-1}(X_{t+1-k} - a_1 X_{t-k} - \cdots - a_p X_{t+1-k-p}).$$

3.2 Asymptotic Normality of Rank Statistics

3.2.1 Asymptotic Distribution of Likelihood Ratios

Under assumptions (A1)-(A4), we have the following results:

Proposition 3.2.1. *Under H_0,*

$$\log L_n(X_1, \ldots, X_n) = \mathcal{L}_n^0(X_1, \ldots, X_n) - \frac{d^2}{2} + 0_p,$$

where

$$\mathcal{L}_n^0(\boldsymbol{X}) = n^{-1/2} \sum_{t=p+1}^{n} \phi(X_t) \sum_{i=1}^{p} d_i X_{t-i}$$

$$d_i = \begin{cases} a_i + b_i & 1 \le i \le \min(p_1, p_2) \\ a_i & p_2 < i \le p_1 \qquad \text{if } p_2 < p_1 \\ b_i & p_1 < i \le p_2 \qquad \text{if } p_1 < p_2 \end{cases}$$

$$p = \max(p_1, p_2) \quad \text{and} \quad d^2 = \textstyle\sum_{i=1}^{p} d_i^2 \sigma^2 I(f).$$

Moreover, $\mathcal{L}_n^0 \xrightarrow{d} \mathcal{N}(0, d^2)$.

The form of this asymptotic distribution shows that, for n sufficiently large, there will be little difference, from a statistical point of view, between AR, MA and ARMA models!!

3.2.2 Asymptotic Distribution of Linear Serial Rank Statistics

In the paper, the authors proposed the linear serial rank statistics for the models as follows:

$$S_n = \frac{1}{n-p} \sum_{t=p+1}^{n} a_n(R_t^{(n)}, R_{t-1}^{(n)}, \ldots, R_{t-p}^{(n)}),$$

$$m_n = E[S_n | H_0^{(n)}] = \frac{1}{n(n-1)\cdots(n-p)} \sum_{1 \le i_1 \ne \cdots \ne i_{p+1} \le n} a_n(i_1, \cdots, i_{p+1}).$$

The authors established the asymptotic equivalence of $(n-p)^{1/2}(S_n - m_n)$ with $\mathcal{S}_n - \mathcal{E}_n$, where

$$\mathcal{S}_n(\boldsymbol{X}) = (n-p)^{-1/2} \sum_{t=p+1}^{n} J(F(X_t), F(X_{t-1}), \ldots, F(X_{t-p}))$$

$$\mathcal{E}_n(\boldsymbol{X}) = \frac{(n-p)^{1/2}}{n(n-1)\cdots(n-p)} \sum_{1 \le t_1 \ne \cdots \ne t_{p+1} \le n} J(F(X_{t_1}), \ldots, F(X_{t_{p+1}})).$$

It is also established that $n^{-1/2}(\mathcal{S}_n - \mathcal{E}_n)$ and $n^{-1/2}\mathcal{L}_n^0$ are asymptotically equivalent to U-statistics.

Proposition 3.2.2. *Under H_0,*

$$\begin{pmatrix} \sqrt{n}(S_n - m_n) \\ \log L_n \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\sum_{i=1}^{p} d_i^2 \sigma^2 I(f) \end{pmatrix}, \begin{pmatrix} V^2 & \sum_{i=1}^{p} d_i C_i \\ \sum_{i=1}^{p} d_i C_i & \sum_{i=1}^{p} d_i \sigma^2 I(f) \end{pmatrix} \right),$$

where

$$\begin{aligned} V^2 &= \int_{[0,1]^{p+1}} [\, J^*(v_{p+1}, \cdots, v_1)\,]^2 dv_1 \cdots dv_{p+1} \\ &\quad + 2\sum_{j=1}^{p} \int_{[0,1]^{p+1+j}} J^*(v_{p+1}, \cdots, v_1) J^*(v_{p+1+j}, \cdots, v_{1+j}) dv_1 \cdots dv_{p+1+j} \end{aligned} \tag{3.2.1}$$

and

$$C_i = \int_{[0,1]^{p+1}} J^*(v_{p+1}, \cdots, v_1) \sum_{j=0}^{p-i} \phi(F^{-1}(v_{p+1-j})) F^{-1}(v_{p+1-j-i}) dv_1 \cdots dv_{p+1}$$

Proposition 3.2.3. *Under H_1,*

$$\sqrt{n}(S_n - m_n) \xrightarrow{d} \mathcal{N}(\sum_{i=1}^{p} d_i C_i, V^2).$$

3.2.3 Asymptotic Efficiency of Linear Serial Rank Statistics

Proposition 3.2.4. *An asymptotically optimal linear serial rank test for H_0 against H_d is provided by any statistic S_n^d with score-generating function (up to additive* *and multiplicative constants) given by*

$$J^d(v_{p+1},\cdots,v_1)=\sum_{i=1}^{p}\frac{d_i}{p+1-i}\sum_{j=0}^{p-i}\phi(F^{-1}(v_{p+1-j}))F^{-1}(v_{p+1-j-i}).$$

Under $H_h(h\in\mathbb{R}^p)$,

$$n^{1/2}(S_n^d-m_n^d)\xrightarrow{d}\mathcal{N}(\sum_{i=1}^{p}h_id_i\sigma^2I(f),V_d^2),$$

where $V_d^2=\sum_{i=1}^{p}d_i^2\sigma^2I(f)$.

This optimality result relies on the following lemma.

Lemma 3.2.5. *Let S_n be a linear rank statistic with score-generating function $J^*(v_{p+1},\cdots,v_1)$, and let*

$$J_0^*(v_{p+1},\cdots,v_1)=(\sigma^2I(f))^{-1}\sum_{i=1}^{p}\frac{C_i}{p+1-i}\sum_{j=0}^{p-i}\phi(F^{-1}(v_{p+1-j}))F^{-1}(v_{p+1-j-i}).$$

Denote by S_n^0 a linear serial rank statistic associated with J_0^*. Then $e(S_n,S_n^0)\le 1$ for any alternative H_n^d.

Proposition 3.2.6. *Under H_0,*

$$\begin{pmatrix}\sqrt{n}r_1\\ \vdots\\ \sqrt{n}r_p\\ \log L_n\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\begin{pmatrix}0\\ \vdots\\ 0\\ -\frac{1}{2}\sum_{i=1}^{p}d_i^2\sigma^2I(f)\end{pmatrix},\begin{pmatrix} & & & d_1\\ & I & & \vdots\\ & & & d_p\\ d_1 & \dots & d_p & \sum d_i^2\sigma^2I(f)\end{pmatrix}\right).$$

Corollary 3.2.7. Under H_d,

$$n^{1/2}\sum_{k=1}^{p}\alpha_kr_k\xrightarrow{d}\mathcal{N}(\sum_{i=1}^{p}\alpha_id_i,\sum_{i=1}^{p}\alpha_i^2)$$

3.3 Appendix

3.3.1 Formulas

$$\beta_s+b_1\beta_{s-1}+\cdots+b_q\beta_{s-q}=0\quad\text{for }\forall s\ge 1.$$

$$(1 + b_1 L + \cdots + b_q L^q)^{-1} = \sum_{k=0}^{\infty} \beta_k L^k.$$

$$e_j = \sum_{k=1}^{j} \beta_{k-1} \left(- \sum_{i=0}^{p} a_i X_{j+1-k-i} \right) + \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k} b_k \right).$$

$$\sum_{k=j+1}^{\infty} \beta_{k-1} \left(\sum_{i=0}^{q} b_i e_{j+1-k-i} \right) = \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k} b_k \right).$$

$$\sum_{k=1}^{j} \beta_{k-1} \left(\sum_{i=0}^{q} b_i e_{j+1-k-i} \right) = e_{j+1-k} - \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k} b_k \right).$$

$$\sum_{i=0}^{p} a_i^0 X_{t-i} = \sum_{i=0}^{q} b_i^0 e_{t-i}(\theta_0).$$

Appendix A: Mathematics amd Time Series

A.1 The Inverse of Partitioned Matrices

For generality, write matrix F as

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}.$$

The inverse of partitioned matrices is shown as

$$F^{-1} = \begin{pmatrix} F_{11}^{-1} + F_{11}^{-1} F_{12} F_{22\cdot1}^{-1} F_{21} F_{11}^{-1} & -F_{11}^{-1} F_{12} F_{22\cdot1}^{-1} \\ -F_{22\cdot1}^{-1} F_{21} F_{11}^{-1} & F_{22\cdot1}^{-1} \end{pmatrix},$$

where

$$F_{22\cdot1} = F_{22} - F_{21} F_{11}^{-1} F_{12}.$$

As an exercise, write matrix G as

$$G = \begin{pmatrix} F_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

$$\begin{aligned}
(F^{-1}-G)F(F^{-1}-G) &= (F^{-1}-G)\begin{pmatrix} 0 & 0 \\ -F_{21}F_{11}^{-1} & I \end{pmatrix} \\
&= \begin{pmatrix} F_{11}^{-1}F_{12}F_{22\cdot 1}^{-1}F_{21}F_{11}^{-1} & -F_{11}^{-1}F_{12}F_{22\cdot 1}^{-1} \\ -F_{22\cdot 1}^{-1}F_{21}F_{11}^{-1} & F_{22\cdot 1}^{-1} \end{pmatrix}\begin{pmatrix} 0 & 0 \\ -F_{21}F_{11}^{-1} & I \end{pmatrix} \\
&= \begin{pmatrix} F_{11}^{-1}F_{12}F_{22\cdot 1}^{-1}F_{21}F_{11}^{-1} & -F_{11}^{-1}F_{12}F_{22\cdot 1}^{-1} \\ -F_{22\cdot 1}^{-1}F_{21}F_{11}^{-1} & F_{22\cdot 1}^{-1} \end{pmatrix} \\
&= F^{-1}-G
\end{aligned}$$

A.2 Discrete Functions and Continuous Functions

Theorem A.1. *If f is a real-valued function defined on $[0,\infty)$ such that $\sup\{V_0^n f := 1,2,\dots\} < \infty$, then $\sum_{k=1}^{\infty} f(k)$ and $\int_0^{\infty} f(t)\,dt$ converge or diverge together.*

note. Of course, the spectral density and periodogram are real-valued functions defined on $[0,\infty)$. It means that $\boldsymbol{m}$ we defined in the paper is also real-valued function. What does V mean? Let me check!

Theorem A.2. *Let f be a nonnegative function defined on $[0,\infty)$. Then $\sum_{k=1}^{\infty} f(k)$ and $\int_0^{\infty} f(t)\,dt$ converge or diverge together provided*

$$\sup\{V_0^n f : n = 1,2,\dots\} < \infty,$$

where $V_0^n f$ denotes the total variation of f on $[0,n]$

Corollary A.3 (Pólya p.37)**.** If a function g ha finite total variation V on $[0,1]$, then

$$\left| \int_0^1 g(x)\,dx - \frac{1}{n}\sum_{k=1}^{n} g(\frac{k}{n}) \right| \le \frac{V}{n}.$$

Calculation. If f is a function satisfying condition (5),

$$\sup\{V_0^n f : n = 1,2,\dots\} < \infty,$$

then we have

$$\begin{aligned}
\int_0^n f(t)\,dt - \sum_{k=1}^{n} f(k) &= n\int_0^1 f(nx)\,dx - \sum_{k=1}^{n} f(k) \\
&= n\left[\int_0^1 g_n(x)\,dx - \frac{1}{n}\sum_{k=1}^{n} g_n\left(\frac{k}{n}\right) \right],
\end{aligned}$$

where $g_n(x)$ is, by definition, equal to $f(nx)$ for all x in $[0,1]$.

Theorem A.4 (Hardy(1910))**.** *Let f be a nonnegative function which is defined and has a continuous derivative on $[0,\infty]$. Then $\sum_{k=1}^{\infty} f(k)$ and $\int_0^{\infty} f(t)\,dt$ converge or diverge together provided*

$$\int_0^{\infty} |f'(t)|\,dt < \infty.$$

Other Results

(3)

$$\sum_{k=1}^{\infty} \sup\{|f(k) - f(t)| : k-1 \leq t \leq k\} < \infty.$$

(6)

$$|\sum_{k=1}^{n} f(k) - \int_0^{n} f(t)\,dt| \leq V_0^n f \quad n = 1, 2, \ldots.$$

(7)

$$\int_0^{\infty} f(t)\,dt - \sum_{k=1}^{n} f(k).$$

A.3 Hölder's inequality

Let (S, Σ, μ) be a measure space and let $1 \leq p, q \leq \infty$ with $p^{-1} + q^{-1} = 1$. Then for all measurable functions f and g on S,

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{A.1}$$

A.4 Uniformly Integrability

Definition A.5. A sequence of random variables $\{X_n\}$ is said to be uniformly integrable if

$$\lim_{\lambda\to\infty} \sup_{n\geq 1} E_n(|X_n| 1_{\{|X_n|\geq\lambda\}}) = 0.$$

The necessary and sufficient condition for uniformly integrability is given as

Lemma A.6. *$\{X_n\}$ is uniformly integrable if and only if*

1. $\sup_{n\geq 1} E_n|X_n| < \infty$*;*
2. *For any sequence of sets $\{B_n\}$ with $B_n \in \mathcal{A}_n$,*

$$P_n(B_n) \to 0 \Rightarrow E_{P_n}(|X_n||1_{\{B_n\}}) \to 0.$$

Remark A.7. Uniform integrability in conjunction with convergence in distribution impies convergence of moments.

A.5 Example of Residual Theorem

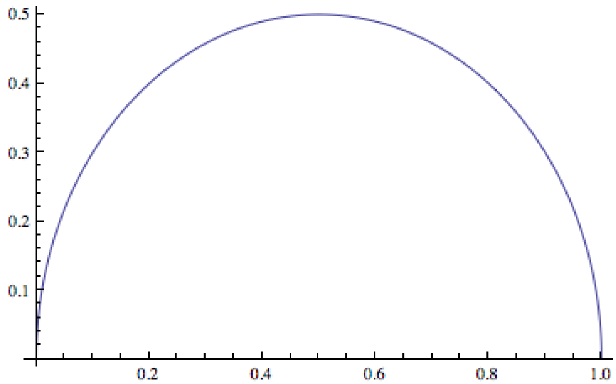
Calculation

$$\int_0^1 \frac{1}{x+1}\,dx = \quad ? \tag{A.2}$$

As what we are taught in the calculus course, you can easily answer this question. The answer is

$$\int_0^1 \frac{1}{x+1}\,dx = [\,\log(x+1)\,]_0^1 = \log 2. \tag{A.3}$$

What we will do here is to apply the residual theorem to it! Let us consider the integral is on the complex plane, and we choose the integral path as follows.



Then,

$$\begin{aligned}\int_0^1 \frac{1}{x+1}\,dx &= -\int_0^\pi \frac{1/2ie^{i\theta}}{1+1/2+1/2e^{i\theta}}\,d\theta = -\int_0^\pi \frac{ie^{i\theta}}{3+1e^{i\theta}}\,d\theta \\ &= \int_{-1}^1 \frac{1}{3+y}\,dy = [\,\log(3+y)\,]_{-1}^1 = \log 2. \end{aligned} \tag{A.4}$$

The most important formula for calculating the integration of spectral density is Formula 1.1 in the following.

Formula 1.1

$$\int_0^{2\pi} \frac{d\theta}{a+b\cos\theta} = \frac{2\pi}{\sqrt{a^2-b^2}}. \tag{A.5}$$

To derive the formula, you only have to think about the singularity in it.

A.6 Time Series Model

We will give some examples of integration of spectrums to look at how powerful residue theorem is in time series.

MA(1)

$$\begin{aligned}\int_{-\pi}^{\pi}(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})\,d\lambda &= \int_{|z|=1}(1-\theta z)(1-\theta/z)\,\frac{dz}{iz}\\ &= \int_{|z|=1}\frac{(1-\theta z)(z-\theta)}{iz^2}\,dz\\ &= 2\pi i\cdot\mathrm{Res}(f_{\mathrm{MA}(1)},0)=2\pi(1+\theta^2).\end{aligned}$$

Formula 2.1

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})\,d\lambda=\sigma^2(1+\theta^2) \tag{A.6}$$

A.7 AR(1)

$$\begin{aligned}\int_{-\pi}^{\pi}\frac{1}{(1-\theta e^{i\lambda})(1-\theta e^{-i\lambda})}\,d\lambda &= \int_{|z|=1}\frac{1}{(1+\theta z)(1+\theta/z)}\,\frac{dz}{iz}\\ &= \int_{|z|=1}\frac{1}{(1+\theta z)(z+\theta)}\,\frac{dz}{i}\\ &= 2\pi i\cdot\mathrm{Res}(f_{\mathrm{AR}(1)},-\theta)=\frac{2\pi}{1-\theta^2}.\end{aligned}$$

Formula 2.2

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{(1-\theta e^{i\lambda})(1-\theta e^{-i\lambda})}\,d\lambda=\frac{\sigma^2}{1-\theta^2} \tag{A.7}$$

A.8 ARMA(1,1)

$$\begin{aligned}
\int_{-\pi}^{\pi} \frac{(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})}{(1-\phi e^{i\lambda})(1-\phi e^{-i\lambda})}\, d\lambda &= \int_{|z|=1} \frac{(1-\theta z)(1-\theta/z)}{(1+\phi z)(1+\phi/z)}\, \frac{dz}{iz} \\
&= \int_{|z|=1} \frac{(1-\theta z)(z-\theta)}{(1+\phi z)(z+\phi)z}\, \frac{dz}{i} \\
&= 2\pi i(\mathrm{Res}(f_{\mathrm{ARMA}(1)}, -\phi) + \mathrm{Res}(f_{\mathrm{ARMA}(1)}, 0)) \\
&= 2\pi \frac{1+2\theta\phi+\theta^2}{1-\phi^2}.
\end{aligned}$$

Formula 2.3

$$\frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{(1+\theta e^{i\lambda})(1+\theta e^{-i\lambda})}{(1-\phi e^{i\lambda})(1-\phi e^{-i\lambda})}\, d\lambda = \sigma^2 \frac{1+2\theta\phi+\theta^2}{1-\phi^2}. \tag{A.8}$$

A.9 AR(2)

$$\begin{aligned}
&\int_{-\pi}^{\pi} \frac{1}{(1-\theta_1 e^{i\lambda}-\theta_2 e^{i2\lambda})(1-\theta_1 e^{-i\lambda}-\theta_2 e^{-i2\lambda})}\, d\lambda \\
=\ &\int_{|z|=1} \frac{z}{(1+\theta_1 z-\theta_2 z^2)(z^2+\theta_1 z-\theta_2)}\, \frac{dz}{i}
\end{aligned}$$

Note that the roots $z_\pm$ of $z^2+\theta_1 z-\theta_2 = 0$ lies in the unit circle, and $z_\pm^2 = -\theta_1 z_\pm + \theta_2$, then

$$1+\theta_1 z_- - \theta_2 z_-^2 = (1-\theta_2^2) + (1+\theta_2)\theta_1 z_- = (1+\theta_2)(1-\theta_2+\theta_1 z_-). \tag{A.9}$$

Note again that $z_+ + z_- = -\theta_1$ and $z_+ z_- = -\theta_2$,

$$\begin{aligned}
\text{(equation above)} &= 2\pi(\mathrm{Res}(f_{\mathrm{AR}(2)}, z_+) + \mathrm{Res}(f_{\mathrm{AR}(2)}, z_-)) \\
&= 2\pi\Big(\frac{z_+}{(1-\theta_2)(1-\theta_2+\theta_1 z_+)(z_+-z_-)} + \frac{z_-}{(1-\theta_2)(1-\theta_2+\theta_1 z_-)(z_- - z_+)}\Big) \\
&= 2\pi \frac{1-\theta_2}{(1+\theta_2)\,[\,(1-\theta_2)^2-\theta_1^2\,]}
\end{aligned}$$

Formula 2.4

$$\frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{(1-\theta_1 e^{i\lambda}-\theta_2 e^{i2\lambda})(1-\theta_1 e^{-i\lambda}-\theta_2 e^{-i2\lambda})}\, d\lambda = \frac{(1-\theta_2)\sigma^2}{(1+\theta_2)\,[\,(1-\theta_2)^2-\theta_1^2\,]} \tag{A.10}$$

A.10 Estimation to the Model with Infinite Variance

Fit the model

$$\beta(L)X_t = \alpha(L)Z_t, \quad t \in \mathbb{Z},$$

where $\beta(L) = 1 - \beta_1 L - \cdots , \beta_p L^p$, $\alpha(L) = 1 + \alpha_1 L + \cdots + \alpha L^q$, and L is backshift operator. Here, the process is not assumed to have an ARMA(p,q) representation.

Denote the parameter

$$\theta = (\beta_1, \ldots, \beta_p, \alpha_1, \ldots, \alpha_q)'.$$

The natural space which contained the parameter is

$$\Theta = \{\theta \in \mathbb{R}^{p+q} : \beta_p \neq 0, \alpha_q \neq 0, \beta(z) \text{ and } \alpha(z) \text{ have no common zeros}, \alpha(z)\beta(z) \neq 0 \text{ for } |z| \leq 1\}.$$

Let $g(\lambda, \theta)$ be the power transfer function

$$g(\lambda, \theta) = \left| \frac{\alpha(\lambda)}{\beta(\lambda)} \right|^2,$$

and $\tilde{I}_{n,X}$ the self-normalized periodogram

$$\tilde{I}_{n,X}(\lambda) = \frac{|\sum_{t=1}^n X_t e^{-i\lambda t}|^2}{\sum_{t=1}^n X_t^2}, \quad -\pi \leq \lambda \leq \pi.$$

In K&M(1995, 1996), the true parameter is defined as θ_0, and if $\theta_0 \in \Theta$, then the estimator

$$\theta_n = \arg\min_{\theta \in \Theta} \sigma_n^2(\theta)$$

is consistent, where

$$\sigma_n^2(\theta) = \int_{-\pi}^{\pi} \frac{\tilde{I}_{n,X}(\lambda)}{g(\lambda, \theta)} d\lambda.$$

Theorem A.8. *Suppose that $(X_t)_{t \in \mathbb{Z}}$ is a general linear process* (2.1.5) *and assumptions (A) and (B) hold. Moreover, assume that the function*

$$h(\theta) = \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda, \theta)} d\lambda$$

has an absolute minimum θ_0 in the parameter space Θ. Then

$$\theta_n \xrightarrow{p} \theta_0$$

and

$$\sigma_n^2(\theta_n) \xrightarrow{p} 2\pi\psi^{-2} h(\theta_0),$$

as $n \to \infty$.

In the same paper, K&M gave the more general theorem as follows.

Theorem A.9. *Suppose that* $(X_t)_{t\in\mathbb{Z}}$ *is a general linear process with representation* (2.1.5) *and* $(Z_t)_{t\in\mathbb{Z}}$ *are i.i.d. symmetric such that*

$$n^{-1/\alpha}\sum_{t=1}^{n} Z_t \xrightarrow{d} Y, \quad (Y \textit{ is symmetric } \alpha\textit{-stable})$$

holds for some $\alpha < 2$. *Moreover, suppose that the function* h *has an absolute minimum* $\theta_0 \in \Theta$. *Then*

$$\left(\frac{n}{\log n}\right)^{1/\alpha}(\theta_n - \theta_0) \xrightarrow{d} 4\pi W^{-1}(\theta_0)\frac{1}{Y_0}\sum_{k=1}^{\infty} Y_k b_k,$$

where $Y_0, Y_1, \ldots$ *are independent random variables,* $Y_0 =_d S_{\alpha/2}(C_{\alpha/2}^{-2/\alpha}, 1, 0)$ *is positive* $\alpha/2$*-stable,* $(Y_t)_{t\in\mathbb{N}}$ *are i.i.d.* $s\alpha s$ *with scale parameter* $\sigma = C_\alpha^{1/\alpha}$, $W^{-1}(\theta_0)$ *is the inverse of the matrix*

$$W(\theta_0) = \int_{-\pi}^{\pi} f(\lambda)\frac{\partial^2 g^{-1}(\lambda,\theta_0)}{\partial\theta^2}\,d\lambda,$$

and, for $k \in \mathbb{N}$, b_k *is the vector*

$$b_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\lambda} f(\lambda)\frac{\partial g^{-1}(\lambda,\theta_0)}{\partial\theta}\,d\lambda,$$

where g^{-1} *denotes the reciprocal of* g.

In the case in which (X_t) *is an ARMA process with true parameter vector* θ_0, *the matrix* $W(\theta_0)$ *can be written in the standard form*

$$W(\theta_0) = \int_{-\pi}^{\pi}\left[\frac{\partial \log g(\lambda,\theta_0)}{\partial\theta}\right]\left[\frac{\partial \log g(\lambda,\theta_0)}{\partial\theta}\right]' d\lambda,$$

which is common for the Whittle estimator with a Gaussian limit.

A.11 Generalized Theorems in Linear Models

Here, the model is vector-valued and is represented by

$$\begin{cases} \underset{(s*1)}{z(n)} = \sum_{j=0}^{\infty} \underset{(s*p)}{G(j)}\underset{(p*1)}{e(n-j)}, \quad n\in\mathbb{Z} \\ E\{e(n)\} = 0 \\ E\{e(m)e(n)'\} = \delta(m,n)K \end{cases} \tag{A.11}$$

Assumption A.10.

$$\sum_{j=0}^{\infty} \mathrm{tr} G(j) K G(j)' < \infty. \tag{A.12}$$

Under this assumption, the process $\{z(n)\}$ is a second-order stationary process. The spectral density matrix of the process is shown as

$$f(\omega) = \frac{1}{2\pi} k(\omega) K k(\omega)^*, \quad -\pi \leq \omega \leq \pi. \tag{A.13}$$

Theorem A.11 (Hosoya-Taniguchi(1982) Thm 2.1)**.** $\{x(t)\}$*: zero-mean second-order stationary process.* $\mathcal{F}_t \equiv \mathcal{F}_t^x$*.*
Assumptions:

1. $\forall \epsilon > 0, \mathrm{Var}\{E(x_\alpha(t+\tau)|\mathcal{F}_t)\} = O(\tau^{-2-\epsilon})$ *uniformly in t, for* $\alpha = 1, \ldots, p$*.*
2. $\forall l, m > t$, $\forall \eta > 0$,
   $E|E\{x_\alpha(l)x_\beta(m)|\mathcal{F}_t)\} - E\{x_\alpha(l)x_\beta(m)\}| = O[\{\min(|l-t|, |m-t|)\}^{-1-\eta}]$ *uniformly in t, for* $\alpha = 1, \ldots, p$*.*
3. *Any element of* $f(\omega) = \{f_{\alpha\beta}(\omega); \alpha, \beta = 1, \ldots, p\}$ *is continuous at the origin; f(0) is non-degenerate.*

Result: $\xi_N = N^{-\frac{1}{2}} \sum_{n=1}^{N} x(n) \to N(0, 2\pi f(0))$.

Theorem A.12 ((HT) Thm 2.2)**.** $\mathcal{B} \equiv \mathcal{B}_e(t)$*.*
Assumptions:

1. $\forall \beta_1, \beta_2, m, \forall \epsilon > 0$,
   $\mathrm{Var}[E\{e_{\beta_1}(n)e_{\beta_2}(n+m)|\mathcal{B}(n-\tau)\} - \delta(m,0)K_{\beta_1\beta_2}] = O(\tau^{-2-\epsilon})$ *uniformly in n.*
2. $\forall \eta > 0$,
   $E|E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)|\mathcal{B}(n_1-\tau)\} - E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)\}| = O(\tau^{-1-\eta})$*, uniformly in* n_1*, where* $n_1 \leq n_2 \leq n_3 \leq n_4$*.*
3. $f_{\beta\beta}$ *are square-integrable.*
4. $\sum_{j_1, j_2, j_3 = -\infty}^{\infty} |Q^e_{\beta_1 \ldots \beta_4}(j_1, j_2, j_3)| < \infty$.

Results:

1. $\sqrt{N}\{C^Z_{\alpha_1\alpha_2}(m) - \gamma^Z_{\alpha_1\alpha_2}(m)\} \to N(0, ...)$

2.

$$\begin{aligned}
&\operatorname{Cov}(\sqrt{N}\{C^Z_{\alpha_1\alpha_2}(m_1)-\gamma^Z_{\alpha_1\alpha_2}(m_1)\},\sqrt{N}\{C^Z_{\alpha_3\alpha_4}(m_2)-\gamma^Z_{\alpha_3\alpha_4}(m_2)\})\\
&\to 2\pi\int_{-\pi}^{\pi}[f_{\alpha_1\alpha_3}(\omega)\bar{f}_{\alpha_2\alpha_4}(\omega)\exp\{-i(m_2-m_1)\omega\}+f_{\alpha_1\alpha_4}(\omega)\bar{f}_{\alpha_2\alpha_3}(\omega)\exp\{i(m_1+m_2)\omega\}]\,d\omega\\
&+2\pi\sum_{\beta_1,\dots,\beta_4=1}^{p}\int\int_{-\pi}^{\pi}\exp\{im_1\omega_1+im_2\omega_2)\}k_{\alpha_1\beta_1}(\omega_1)\\
&\quad k_{\alpha_2\beta_2}(-\omega_1)k_{\alpha_3\beta_3}(\omega_2)k_{\alpha_4\beta_4}(-\omega_2)\tilde{Q}^e_{\beta_1\dots\beta_4}(\omega_1,-\omega_2,\omega_2)\,d\omega_1 d\omega_2
\end{aligned}$$

Lemma A.13 ((HT) Lem 3.1)**.** *$D(f_{T(f)},f)=\min_{t\in\Theta}D(f_t,f)$.*
Assumptions:

1. *$\Theta:(\subset\mathbb{R}^q)$ compact.*
2. *$\theta_1\neq\theta_2\Rightarrow f_{\theta_1}\neq f_{theta_2}$.*
3. *$f_\theta(\omega)$: positive definite.*
4. *$f_\theta(\omega)$ is continuous w.r.t θ,ω.*

Results:

1. *$\forall f\in\mathcal{P},\exists T(f)\in\Theta$ s.t. $D(f_{T(f)},f)=\min_{t\in\Theta}D(f_t,f)$.*
2. *T(f): unique, $T(f_N)\to_\omega f\implies$ as $N\to\infty,T(f_N)\to_\omega f$.*
3. *$\forall\theta\in\Theta,T(f_\theta)=\theta$.*

Theorem A.14 ((HT) Thm 3.1)**.** *$\exists 1\,T(f)\in\Theta^\circ$;*

$$M_f=\int_{-\pi}^{\pi}\left[\frac{\partial^2}{\partial\theta\partial\theta'}\operatorname{tr}\{f_t(\omega)^{-1}f(\omega)\}+\frac{\partial^2}{\partial\theta\partial\theta'}\log\det f_\theta(\omega)\right]_{\theta=T(f)}d\omega,$$

where M_f is nonsingular matrix.
Assumptions:

1. *$\forall\beta_1,\beta_2,m,\forall\epsilon>0$,*
   $\operatorname{Var}[E\{e_{\beta_1}(n)e_{\beta_2}(n+m)|\mathcal{B}(n-\tau)\}-\delta(m,0)K_{\beta_1\beta_2}]=O(\tau^{-2-\epsilon})$ uniformly in n.
2. *$\forall\eta>0$,*
   $E|E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)|\mathcal{B}(n_1-\tau)\}-E\{e_{\beta_1}(n_1)e_{\beta_2}(n_2)e_{beta_3}(n_3)e_{\beta_4}(n_4)\}|=O(\tau^{-1-\eta})$, uniformly in n_1, where $n_1\le n_2\le n_3\le n_4$.

3. $f_{\beta\beta}$ *are square-integrable.*

4. $\sum_{j_1,j_2,j_3=infty}^{\infty} |Q^e_{\mathcal{B}_1\ldots\mathcal{B}_4}(j_1,j_2,j_3)| < \infty$;

5. $f(\omega) \in \mathrm{Lip}(\alpha)$ *where* $\alpha > \frac{1}{2}$.

Results:

1. $p\text{-}lim_{N\to\infty} T(I_z) = T(f)$.

2. as $N \to \infty$, $\sqrt{N}\{T(I_z) - T(f)\} \to N(0, M_f^{-1}\tilde{V}M_f^{-1})$. *where*

$$\tilde{V}_{jl} = 4\pi \int_{-\pi}^{\pi} \mathrm{tr}\left[f(\omega)\frac{\partial}{\partial\theta_j}\{f_t(\omega)\}^{-1} f(\omega)\frac{\partial}{\partial\theta_l}\{f_t(\omega)\}^{-1}\right]_{\theta=T(f)} d\omega$$

$$+2\pi \sum_{r,t,u,v=1}^{s} \int\int_{-\pi}^{\pi} \left\{\frac{\partial}{\partial\theta_j} f_\theta^{(r,t)}(\omega_1)\frac{\partial}{\partial\theta_l} f_\theta^{(u,v)}(\omega_2)\right\}_{\theta=T(f)} \tilde{Q}^z_{rtuv}(-\omega_1,\omega_2,-\omega_2)\, d\omega_1 d\omega_2,$$

where $f_\theta^{(r,t)}(\omega)$ *is the (r,t) element of* $\{f_\theta(\omega)\}^{-1}$.

Corollary A.15 ((HT)Cor 3.1)**.**

$$\tilde{V}_{jl} = 4\pi \int_{-\pi}^{\pi} \mathrm{tr}\left[f(\omega)\frac{\partial}{\partial\theta_j}\{f_t(\omega)\}^{-1} f(\omega)\frac{\partial}{\partial\theta_l}\{f_t(\omega)\}^{-1}\right]_{\theta=T(f)} d\omega$$
$$+ 2\pi \sum_{a,b,c,d=1}^{p} \sum_{r,t,u,v=1}^{s} \int\int_{-\pi}^{\pi} \left\{\frac{\partial}{\partial\theta_j} f_\theta^{(r,t)}(\omega_1)\frac{\partial}{\partial\theta_l} f_\theta^{(u,v)}(\omega_2)\right\}_{\theta=T(f)}$$
$$k_{ra}(-\omega_1)k_{tb}(\omega_1)k_{uc}(-\omega_2)k_{vd}(\omega_2)\tilde{Q}^e_{abcd}(-\omega_1,\omega_2,-\omega_2)\, d\omega_1 d\omega_2$$

Proposition A.16 ((HT) Prop 3.1)**.**
Assumption:

$$\mathrm{cum}\{e_a(n_1), e_b(n_2), e_c(n_3), e_d(n_4)\} = \begin{cases} \kappa_{abcd} & \text{if } n_1 = n_2 = n_3 = n_4 \\ 0 & \text{otherwise.} \end{cases} \tag{A.14}$$

Result:

$$\tilde{V}_{jl} = 4\pi \int_{-\pi}^{\pi} \mathrm{tr}\left[f(\omega)\frac{\partial}{\partial\theta_j}\{f_t(\omega)\}^{-1} f(\omega)\frac{\partial}{\partial\theta_l}\{f_t(\omega)\}^{-1}\right]_{\theta=T(f)} d\omega$$
$$+ \sum_{a,b,c,d=1}^{s} \kappa_{abcd} \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} k^*(\omega)\frac{\partial}{\partial\theta_j}\{f_\theta(\omega)\}^{-1}k(\omega)d\omega\right]_{ab}$$
$$\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} k^*(\omega)\frac{\partial}{\partial\theta_l}\{f_\theta(\omega)\}^{-1}k(\omega)d\omega\right]_{cd} \Big|_{\theta=T(f)} .$$

In the case where $f(\omega) = f_\theta(\omega)$ and θ is the innovation-free parameter, the second term in the right-hand side will be 0. On the other hand, in the case of $f(\omega) \neq f_\theta(\omega)$, even if (A.14) is satisfied, the quasi-Gaussian maximum likelihood estimates for the innovation-free parameters are generally not robust against the fourth cumulant. In the case $s = 1$, that is in the scalar case, the quasi-Gaussian maximum likelihood estimates for the innovation-free parameters are robust against fourth cumulant even if $f(\omega) \neq f_\theta(\omega)$.

A.12 Some Results for Unit Root Case

Suppose that $\{Y_t : t = 1, \dots, n\}$ is generated by the first-order autoregressive process

$$Y_t = \theta Y_{t-1} + e_t, \quad Y_0 = 0,\, t = 1, \dots, \tag{A.15}$$

where e_t's are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables and

$$\theta = \exp(\frac{c}{n}).$$

As a generation of the LSE $\hat{\theta}$ and $\hat{\hat{\theta}}_{c_1,c_2}$,

$$\hat{\theta}_{c_1,c_2} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1}}{\sum_{t=2}^{n-1} Y_t^2 + c_1 Y_1^2 + c_2 Y_n^2}, \quad c_1, c_2 \geq 0.$$

$$\hat{\hat{\theta}}_{c_1,c_2} = \frac{\sum_{t=2}^{n} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=2}^{n-1} (Y_t - \bar{Y})^2 + c_1 (Y_1 - \bar{Y})^2 + c_2 (Y_n - \bar{Y})^2}, \quad c_1, c_2 \geq 0,\, \bar{Y} = \sum_{t=1}^{n} Y_t / n,$$

are supposed.

The hypothesis is supposed as

$$H : \theta = 1 \quad vs \quad A : \theta \in (0, 1).$$

For the testing problem, the following tests are introduced:

$$K_{1n} = \frac{\sqrt{2}}{n \hat{\sigma}^2} \sum_{t=2}^{n} (\hat{\theta}_{c_1,c_2} - 1); \tag{A.16}$$

$$K_{2n} = \frac{n}{\sqrt{2}} (\hat{\theta}_{c_1,c_2} - 1); \tag{A.17}$$

$$K_{3n} = \left(\sum_{t=2}^{n} \frac{Y_{t-1}^2}{\hat{\sigma}^2} \right)^{1/2} (\hat{\theta}_{c_1,c_2} - 1), \tag{A.18}$$

where $\hat{\sigma}^2 = n^{-1} \sum_{t=2}^{n} (Y_t - \hat{\theta}_{c_1,c_2} Y_{t-1})^2$.

Ornstein-Uhlenbeck process

Let $J_c(t)$ be an Ornstein-Uhlenbeck process

$$J_c(t) = \int_0^t \exp\{(t-s)c\}\, dW(s),$$

which is generated by

$$dJ_c(t) = cJ_c(t)dt + dW(t),$$

with initial condition $J_c(0) = 0$.

Integrated process

For the process above,
when $c \neq 0$, it is called a near-integrated process;
when $c = 0$, it is called an integrated process.

We give Assumptions on the unit root case:

(UR1) $E(e_t) = 0$ for all t,

(UR2) $\sup_t E|e_t|^{\beta+\epsilon} < \infty$ for some $\beta > 2$ and $\epsilon > 0$,

(UR3) $\sigma^2 = \lim_{n\to\infty} E(n^{-1}S_n^2)$ exists and $\sigma^2 > 0$ where $S_t = \sum_{s=1}^t e_s$,

(UR4) $\{e_t\}$ is strong mixing with mixing coefficients α_m that satisfy $\sum_{m=1}^\infty \alpha_m^{1-2/\beta} < \infty$.

Some famous results are given below.

Lemma A.17 (Phillips(1987b))**.** *If $\{Y_t\}$ is a near-integrated time series generated by* (A.15)*, then, as $n \to \infty$,*

1. $n^{1/2}Y_{[nt]} \xrightarrow{d} \sigma J_c(t)$;
2. $n^{3/2}\sum_{t=1}^n Y_t \xrightarrow{d} \sigma \int_0^1 J_c(t)\, dt$;
3. $n^{-2}\sum_{t=1}^n Y_t^2 \xrightarrow{d} \sigma^2 \int_0^1 J_c(t)^2\, dt$;
4. $n^{-1}\sum_{t=1}^n Y_{t-1}e_t \xrightarrow{d} \sigma^2 \int_0^1 J_c(t)\, dW(t) + \frac{1}{2}(\sigma^2 - \sigma_e^2)$ *with* $\sigma_e^2 = \lim_{n\to\infty} n^{-1}\sum_{t=1}^n E(e_t^2)$.

Theorem A.18. *If $\{Y_t\}$ is a near-integrated time series generated by the model above, then, as $n \to \infty$,*

$$n(\hat{\theta}_{c_1,c_2} - \theta) \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\, dt - \sigma_e^2/\sigma^2}{2\int_0^1 J_c(t)^2\, dt}.$$

Corollary A.19. If $\theta = 1$ (i.e., $c = 0$), then

$$n(\hat{\theta}_{c_1,c_2} - 1) \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - \sigma_e^2/\sigma^2}{2\int_0^1 W(t)^2\,dt}.$$

Theorem A.20. *If $\{Y_t\}$ is a near-integrated time series generated by the model above, then, as $n \to \infty$,*

$$n(\hat{\hat{\theta}} - \theta) \xrightarrow{d} \frac{-2cG + (1-2c_2)T^2 + 4c_2TH - 2(c_1+c_2-1)H^2 - 2HW(1) - \sigma_e^2/\sigma^2}{2(G - H^2)},$$

where $G = \int_0^1 J_c(t)^2\,dt$, $T = J_c(1)$ and $H = \int_0^1 J_c(t)\,dt$.

Corollary A.21. If $\theta = 1$, then, as $n \to \infty$,

$$n(\hat{\hat{\theta}}_{c_1,c_2} - 1) \xrightarrow{d} \frac{(1-2c_2)T_w^2 + 2(2c_2-1)T_wH_w - 2(c_1+c_2-1)H_w^2 - \sigma_e^2/\sigma^2}{2(G_w - H_w^2)},$$

where $G_w = \int_0^1 W(t)^2\,dt$, $T_w = W(1)$ and $H_w = \int_0^1 W(t)\,dt$.

Theorem A.22. *Under H, as $n \to \infty$, we have*

$$K_{1n} \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - 1}{\sqrt{2}}; \tag{A.19}$$

$$K_{2n} \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - 1}{2\sqrt{2}\int_0^1 W(t)^2\,dt}; \tag{A.20}$$

$$K_{3n} \xrightarrow{d} \frac{(1-2c_2)W(1)^2 - 1}{2(\int_0^1 W(t)^2\,dt)^{1/2}}. \tag{A.21}$$

Theorem A.23. *Under A_n, as $n \to \infty$, we have*

$$K_{1n} \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\,dt - 1}{\sqrt{2}}; \tag{A.22}$$

$$K_{2n} \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\,dt - 1}{2\sqrt{2}\int_0^1 J_c(t)^2\,dt}; \tag{A.23}$$

$$K_{3n} \xrightarrow{d} \frac{(1-2c_2)J_c(1)^2 - 2c\int_0^1 J_c(t)^2\,dt - 1}{2(\int_0^1 J_c(t)^2\,dt)^{1/2}}. \tag{A.24}$$

Appendix B: Regularly Varying Tail with Index α

Before introducing the stable random variables, we have to understand the scope where the stable random variables dominate. The crucial concept here is ”in the domain of attraction”.

B.1 Domain of Attraction

ϵ is in the domain of attraction of a stable law with a parameter α and write $\epsilon \in \mathcal{R}_{-\alpha}$ if

$$P(\epsilon > x) = c_1 x^{-\alpha} L(x)(1 + \alpha_1(x)), \quad x > 0, c_1 \geq 0, \tag{B.1}$$

and

$$P(\epsilon < -x) = c_2 x^{-\alpha} L(x)(1 + \alpha_2(x)), \quad x > 0, c_2 \geq 0, \tag{B.2}$$

with $0 < \alpha < 2$, $L(x)$ a slowly varying function at ∞ and $\alpha_i(x) \to 0$ as $|x| \to \infty$. If $L(x) = 1$, then ϵ is in the normal domain of attraction of a stable law with parameter α.

Another form for the random variables in the domain of attraction with distribution function F satisfies

$$\begin{cases} x^\alpha P(\epsilon > x) = x^\alpha(1 - F(x)) \to pC, & x > 0, \\ x^\alpha P(\epsilon < -x) = x^\alpha F(-x) \to qC, & x > 0, \end{cases} \tag{B.3}$$

which means

$$\begin{cases} c_1 = pC, \\ c_2 = qC. \end{cases} \tag{B.4}$$

B.2 Parametrization of stable distributions

Let Y be distributed as stable distribution $S_\alpha(\sigma, \beta, \mu)$, then its characteristic function is

$$E(e^{itY}) = \begin{cases} \exp\{-\sigma^\alpha |t|^\alpha (1 - i\beta(\text{sign}t)\tan\frac{\pi\alpha}{2}) + i\mu t\} & \alpha \neq 1, \\ \exp\{-\sigma|t|(1 + \frac{2i\beta}{\pi}(\text{sign}t)\log|t|) + i\mu t\}, \end{cases} \tag{B.5}$$

where σ is the scale parameter, β is the skewness parameter and μ is the location parameter.

Stable random variables has an exact form of their tails, that is,

$$\begin{cases} x^\alpha P(\epsilon > x) = \frac{1+\beta}{2}\sigma^\alpha C_\alpha, & x > 0, \\ x^\alpha P(\epsilon < -x) = \frac{1-\beta}{2}\sigma^\alpha C_\alpha, & x > 0. \end{cases} \tag{B.6}$$

Here C_α is a constant depending on α, and

$$C_\alpha = \begin{cases} \frac{1}{\Gamma(1-\alpha)\cos(\pi\alpha/2)}, & \text{if } \alpha \neq 1, \\ \frac{2}{\pi} & \text{if } \alpha = 1. \end{cases} \tag{B.7}$$

Since $0 < \alpha < 2$ and $\Gamma(2 - \alpha) = (1 - \alpha)\Gamma(1 - \alpha)$,

$$C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, & \text{if } \alpha \neq 1, \\ \frac{2}{\pi} & \text{if } \alpha = 1. \end{cases} \tag{B.8}$$

is used in some books.

B.3 Some Results

Assume that X and Y are independent and satisfy

$$P[\,|X| > t\,] \in \mathcal{R}_{-\alpha}, \quad p_1 = \lim_{t\to\infty} \frac{P[\,X > t\,]}{P[\,|X| > t\,]} \quad \text{exists}$$

and

$$P[\,|Y| > t\,] \in \mathcal{R}_{-\alpha}, \quad p_1 = \lim_{t\to\infty} \frac{P[\,Y > t\,]}{P[\,|Y| > t\,]} \quad \text{exists,}$$

and $0 < \alpha \leq 2$.

For $\alpha < 2$, these conditions are necessary and sufficient for the distributions of X and Y to be in stable domains of attraction. Define $X_+ = \max(0, X)$, $X_- = \max(0, -X)$ and similarly for Y_+, Y_-.

Proposition B.1. *Under each of the following conditions $p = \lim_{t\to\infty} \frac{P[\,XY>t\,]}{P[\,|XY|>t\,]}$ exists and hence the distribution of XY is in a stable α domain of attraction.*

(B1) $P[\,X > t\,] = p_1 P[\,|X| > t\,]$ *and either* $p_1 = 1/2$ *or* $P[\,Y > t\,] = p_2 P[\,|Y| > t\,]$. *In this case* $p = p_1 p_2 + (1 - p_1)(1 - p_2)$.

(B2) $P[\,|X| > e^t\,]$ *and* $P[\,|Y| > e^t\,]$ *are both in* $\mathcal{S}_\alpha$ *and* $\lim_{t\to\infty} \frac{P[\,|Y|>t\,]}{P[\,|X|>t\,]} = k < \infty$. *In this case*

$$p = \frac{(p_1 E Y_+^\alpha + (1 - p_1) E Y_-^\alpha) + (p_2 E X_+^\alpha + (1 - p_2) E X_-^\alpha)}{E|Y|^\alpha + k E|X|^\alpha}$$

(B3) Either $\frac{P[\,|XY|>t\,]}{P[\,|X|>t\,]} \to E|Y|^\alpha < \infty$ *or* $\frac{P[\,XY>t\,]}{P[\,X>t\,]} \to EY_+^\alpha < \infty$, $p_1 > 0$. *In this case*

$$p = \frac{p_1 E Y_+^\alpha + (1 - p_1) E Y_-^\alpha}{E|Y|^\alpha}.$$

(B4) $E|X|^\alpha = E|Y|^\alpha = \infty$. *In this case* $p = p_1 p_2 + (1 - p_1)(1 - p_2)$.

Remark B.2. It is easy to see from (4) that the product of stable distribution is in stable domain of attraction and another tail satisfies

$$q = 1 - p = p_1(1 - p_2) + p_2(1 - p_1)$$

if you substitute $-t$ for t above.

B.4 Point Process for the case of dependent variables

Let $\mathcal{F}_s$ be the collection of step functions on $\mathbb{R} - \{0\}$ with bounded support. We have the following conditions:

On sequence

(a1) $\{X_k\}$ is an i.i.d sequence of random variables.

(a2) $\{X_k\}$ is a strictly stationary sequence of random variables.

On random variables

The condition (b1) on random variables is:

$$P(|X_k| > x) = x^{-\alpha} L(x) \tag{B.9}$$

with $\alpha \in (0, 2)$ and $L(x)$ a slowly varying function at ∞;

$$\frac{P(X_k > x)}{P(|X_k| > x)} \to p, \quad \frac{P(X_k < -x)}{P(|X_k| > x)} \to q \tag{B.10}$$

as $x \to \infty$, $0 \leq p \leq 1$ and $q = 1 - p$.

On technical conditions

(c1) Let a_n be defined as

$$a_n = \inf \left\{ x : P(|X_1| > x) \leq n^{-1} \right\}. \tag{B.11}$$

(d1) The mixing condition is defined as

$$E \exp\left(-\sum_{j=1}^{n} f(X_j/a_n) \right) - \left(E \exp\left(-\sum_{j=1}^{r_n} f(X_j/a_n) \right) \right)^{[n/r_n]} \to 0 \tag{B.12}$$

as $n \to \infty$ for all $f \in \mathcal{F}_s$.

B.5 Explanation

- (a1) and (b1) are necessary and sufficient for the existence of normalizing constants a_n, b_n for which $(S_n - b_n)/a_n$ converges weakly to some stable law with index α (cf. Feller (1971)). They also imply that

$$\lim_{n \to \infty} \frac{P(S_n > t_n)}{nP(X_1 > t_n)} = 1 \tag{B.13}$$

for any constants t_n satisfying $nP(X_1 > t_n) \to 0$.

- (a1), (b1) and (c1) imply that

$$nP(|X_1| > a_n x) \to x^{-\alpha} \quad \text{for all } x > 0. \tag{B.14}$$

Or we can write it more implicitly,

$$nP(X_1/a_n \in \cdot) \to_v \mu(\cdot), \tag{B.15}$$

where μ is the measure

$$\mu(dx) = \lambda(dx) = \alpha p x^{-\alpha-1} 1_{(0,\infty)}(x)dx + \alpha q(-x)^{-\alpha-1} 1_{(-\infty,0)}(x)dx, \tag{B.16}$$

and $\to_v$ denotes vague convergence on $\mathbb{R} - \{0\}$.

Define the point process

$$N_n = \sum_{j=1}^{n} \delta_{X_j/a_n}, \tag{B.17}$$

where δ_x represents unit point measure at the point x.

For any $y \geq 0$, define

$$M_y = \{\mu \in M : \mu([-y, y]^c) > 0 \text{ and } \mu([-x, x]^c) = 0 \text{ for some } 0 < x(= x_\mu) < \infty\}. \tag{B.18}$$

For $\mu \in M_0$, let $\mu_+ = \max(0, \text{largest point of } \mu)$, $\mu_- = \min(0, \text{smallest point of } \mu)$ and $x_\mu = \max(\mu_+, \mu_-)$. Define a mapping on M_0 by

$$\Omega : \mu \to (x_\mu, \mu(x_\mu \cdot)). \tag{B.19}$$

The mapping Ω is continuous with range $(0, \infty) \times \tilde{M}$, where $\tilde{M} = \{\mu \in M : \mu([-1, 1]^c) = 0,\ \mu(\{-1\} \cup \{1\}) > 0\}$. Denote by $\mathcal{B}(\tilde{M})$ the Borel σ-field of $\tilde{M}$.

$\gamma := \lambda\{\mu : \mu([-1, 1]^c) > 0\} \in (0, 1]$

$$X_n = \sum_{j=0}^{\infty} c_j Z_{n-j}. \tag{B.20}$$

$$\tilde{a}_n = \inf\left\{ x : P(|Z_0 Z_1| > x) \leq n^{-1} \right\}. \tag{B.21}$$

$$\sum_{j=0}^{\infty} |c_j|^\delta < \infty \quad \text{for some} \quad \delta < \alpha,\ \delta \leq 1. \tag{B.22}$$

$$\sum_{j=-\infty}^{\infty} |c_j|^\delta |j| < \infty \quad \text{with} \quad \begin{cases} \delta = 1, & \text{if } \alpha > 1 \\ 0 < \delta < \alpha & \text{if } \alpha \leq 1. \end{cases} \tag{B.23}$$

$$\hat{\rho}(h) = \frac{C(h)}{C(0)}, \quad h \geq 0, \tag{B.24}$$

where

$$C(h) = \sum_{t=1}^{n} X_t X_{t+h}. \tag{B.25}$$

Further,

$$\rho(h) = \frac{\sum_j c_j c_{j+h}}{\sum_j c_j^2}.$$

B.6 Theorems of Asymptotic Distributions

Theorem B.3 (Davis and Hsing [1995]). *Assume that the condition (d1) holds for $\{X_j\}$, and $N_n \to_d$ some $N \neq o$. Then N is infinitely divisible with canonical measure λ satisfying $\lambda(M_0^c) = 0$ and $\lambda \circ \Omega^{-1} = \nu \times \mathcal{O}$, where $\mathcal{O}$ is a probability measure on $(M, \mathcal{B}(\tilde{M}))$, and*

$$\nu(dy) = \gamma \alpha y^{-\alpha-1} I_{(0,\infty)}(y)\, dy. \tag{B.26}$$

In this case the Laplace transform of N is

$$\exp\left\{-\int_0^\infty \int_{\tilde{M}} (1 - \exp(-\mu f(y\cdot)))\mathcal{O}(d\mu)\nu(dy)\right\}, \quad f \in \mathcal{F}. \tag{B.27}$$

Theorem B.4 (Davis and Hsing [1995]). *Under the condition (d1) for $\{X_j\}$, the following are equivalent:*

(i) *N_n converges in distribution to some $N \neq o$.*

(ii) *For some finite positive constant γ, $k_n P[\vee_{k=1}^{r_n} |X_k| > a_n x] \to \gamma x^{-\alpha}$, $x > 0$, and for some probability measure $\mathcal{O}$ on $\tilde{M}$, $P[\sum_{j=1}^{r_n} \delta_{X_j/(\vee_1^{r_n}|X_k|)}] \in \cdot |\vee_1^{r_n} |X_k|| > a_n x] \to_w \mathcal{O}$, $x > 0$.*

In this case N is infinitely divisible with canonical measure λ confined to M_0 and satisfying

$$\lambda \circ \Omega^{-1} = \nu \times \mathcal{O}, \tag{B.28}$$

where $\nu(dy) = \gamma \alpha y^{-\alpha-1}\, dy$.

Theorem B.5 (Davis and Hsing [1995]). *Suppose that $N_n \to_d N$ and N has the representation given by Theorem B.3. Then $\gamma \sum_{i=1}^{\infty} E|Q_i|^\alpha \leq 1$, where $\sum_{i=1}^{\infty} \delta_{Q_i} \sim \mathcal{O}$. The equality holds if $\{N_n([-1,1]^c)\}_{n=1}^{\infty}\}$ is uniformly integrable.*

Theorem B.6 (Davis and Hsing [1995]). *Suppose that* $\{X_j\}$ *is a stationary sequence of random variables for which all finite-dimensional distributions are jointly regularly varying with index* $\alpha > 0$. *To be specific, let* $\boldsymbol{\theta}^{(m)} = (\theta_i^{(m)}, |i| \le m)$ *be the ransom vector* $\boldsymbol{\theta}$ *that appears in the definition of joint regular variation of* X_i, $|i| \le m$. *Assume that the condition (d1) holds for* $\{X_j\}$ *and that*

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left[\bigvee_{m \le |i| \le r_n} |X_i| > ta_n \middle| |X_0| > ta_n \right] = 0, \quad t > 0, \tag{B.29}$$

where a_n *is defined above. Then the limit*

$$\gamma := \lim_{m\to\infty} \frac{E(|\theta_0^{(m)}|^\alpha - \vee_{j=1}^m |\theta_j^{(m)}|^\alpha)_+}{E|\theta_0^{(m)}|^\alpha} \tag{B.30}$$

exists. If $\gamma = 0$ *then* $N_n \to_d o$; *if* $\gamma > 0$, *then* N_n *converges in distribution to some* N, *where, using the representation* $\lambda \circ \Omega^{-1} = \nu \times \mathcal{O}$ *described in Theorem B.3,* $\nu(dy) = \gamma\alpha y^{-\alpha-1}\, dy$ *and* $\mathcal{O}$ *is the weak limit of*

$$\lim_{m\to\infty} \frac{E(|\theta_0^{(m)}|^\alpha - \vee_{j=1}^m |\theta_j^{(m)}|^\alpha)_+ I(\sum_{|i|\le m} \delta_{\theta_i^{(m)}} \in \cdot)}{E(|\theta_0^{(m)}|^\alpha - \vee_{j=1}^m |\theta_j^{(m)}|^\alpha)_+} \tag{B.31}$$

as $m \to \infty$, *which exists.*

Theorem B.7 (Davis and Hsing [1995]). *Let* $\{X_j\}$ *be a strictly stationary sequence satisfying (c1) and*

Theorem B.8 (Davis and Resnick [1986]). *Let* $\{\, Z_t \,\}$ *be iid satisfying* (B.9) *and* (B.10) *with* $0 < \alpha < 2$ *and* $E|Z_1|^\alpha = \infty$. *Then, if* a_n *and* $\tilde{a}_n$ *are given by* (B.11) *and* (B.21),

$$(a_n^{-2} \sum_{t=1}^n Z_t^2, \tilde{a}_n^{-1} \sum_{t=1}^n (Z_t Z_{t+1} - \mu_n), \ldots, \tilde{a}_n^{-1} \sum_{t=1}^n (Z_t Z_{t+h} - \mu_n)) \Rightarrow (S_0, S_1, \ldots, S_h), \tag{B.32}$$

where $\mu_n = E Z_1 Z_2 1_{[\,|Z_1 Z_2| \le \tilde{a}_n\,]}$ *and* $S_0, S_1, \ldots, S_h$ *are independent stable random variables;* S_0 *is positive with index* $\alpha/2$ *and* $S_1, S_2, \ldots, S_h$ *are identically distributed with index* α.

Theorem B.9. *Suppose* $X_t = \sum_{j=-\infty}^{\infty} c_j Z_{t-j}$ *where* $\{c_j\}$ *satisfies* (B.23) *and* $\{Z_t\}$ *satisfies* (B.9) *and* (B.10), *and* $E|Z_1|^\alpha = \infty$, $0 < \alpha < 2$. *If* a_n *and* $\tilde{a}_n$ *are given by* (B.11) *and* (B.21), *then for any positive integer* l,

$$(\tilde{a}_n^{-1} a_n^2 (\hat{\rho}(h) - \rho(h) - d_{h,n}/C(0)), 1 \le h \le l) \Rightarrow (Y_1, Y_2, \ldots, Y_l) \tag{B.33}$$

in $\mathbb{R}^l$, *where*

$$\begin{aligned} d_{h,n} &= \sum_{j=1}^{\infty}(\rho(h+j)+\rho(h-j)-2\rho(h)\rho(h))\sum_i c_i^2 EZ_1Z_2 1_{[|Z_1Z_2|\le \tilde{a}_n]}, \\ Y_h &= \sum_{j=1}^{\infty}(\rho(h+j)+\rho(h-j)-2\rho(j)\rho(h)S_j/S_0), \end{aligned}$$

and $S_0, S_1, S_2, \ldots$ *are independent stable random variables as described in Theorem 3.3. In addition, if either*

case 1 $0 < \alpha < 1$, *or*

case 2 $\alpha = 1$ *and the distribution of* Z_t *is symmetric, or*

case 3 $1 < \alpha < 2$ *and* $EZ_1 = 0$,

then (B.33) *holds with* $d_{h,n} = 0$, $h = 1, \ldots, l$, *and a location change in the* S_j*'s,* $j \geq 1$.

Theorem B.10 (Davis and Resnick [1985]). *Let* $\sum_{k=1}^{\infty} \epsilon_{j_k}$ *be a PRM(λ) on* $\mathbb{R}\backslash\{0\}$ *with*

$$\lambda(dx) = \alpha p x^{-\alpha-1} 1_{(0,\infty)}(x)dx + \alpha q(-x)^{-\alpha-1} 1_{(-\infty,0)}(x)dx, \quad 0 < \alpha < 2. \qquad (B.34)$$

Suppose (B.9) - (B.20), (B.14), (B.22) *hold with* $0 < \alpha < 2$. *Then for every nonnegative integer* l, *as* $n \to \infty$*:*

(DR-1)

$$(n/a_n^2)(\hat{\gamma}(0), \hat{\gamma}(1), \cdots, \hat{\gamma}(l)) \Rightarrow \sum_{i=1}^{\infty} j_i^2 (\sum_{j=0}^{\infty} c_j^2, \sum_{j=0}^{\infty} c_j c_{j+1}, \cdot, \sum_{j=0}^{\infty} c_j c_{j+l}) \qquad (B.35)$$

and

(DR-2)

$$\hat{\rho}(l) \to \rho(l) = \frac{\sum_{j=0}^{\infty} c_j c_{j+l}}{\sum_{j=0}^{\infty} c_j^2} \quad \text{in probability.} \qquad (B.36)$$

Appendix C: Generalized Domain of Attraction

Let X be a real Banach space, that is, X is a real linear, normed, complete space, with norm $\|\cdot\|$. By X^* we denote its *topological dual Banach*, that is, $x^* \in X^*$ are continuous linear functionals on X, and $\langle \cdot, \cdot \rangle$ is the dual pair between X^* and X. When the norm in X is given by a scalar product, X is called a *Hilbert* space. In that case, X^* is isomorphic to X and the dual pair is simply the scalar product. Furthermore, all real separable Hilbert spaces are isomorphic to l_2, the space of all real square-summable sequences with

$$\langle x, y \rangle := \sum_i x_i y_i, \quad \|x\| := \langle x, x \rangle^{1/2}.$$

The collection $L(X, Y)$ of all bounded linear operators from X into Y, using the operator norm, is also Banach space. Here, the assumption that A is bounded and linear is equivalent to A being continuous and linear form X to Y, where the topologies are given by the norms. When $X = Y$, $L(X, Y)$ is denoted by $\mathrm{End}(X)$; in which case, we also have that the product of two operators in $\mathrm{End}(X)$ is a continuous linear operator: if $A, B \in \mathrm{End}(X)$, then $AB : X \to X$ is given by $(AB)x = A(Bx)$ for $x \in X$. Moreover, $\|AB\| \leq \|A\|\|B\|$ for all $A, B \in \mathrm{End}(X)$. With this multiplication of operators, $\mathrm{End}(X)$ becomes a topological semigroup. By $\mathrm{Aut}(X)$, we denote the set of all invertible operators in $\mathrm{End}(X)$. These inverse are also continuous and linear, so $\mathrm{Aut}(X)$ is a topological group.

Theorem C.1. *Let ξ_n, ξ be $\mathbb{R}^d$-valued random variables. Then ξ_n converges in distribution to ξ in $\mathbb{R}^d$ if and only if for every $a \in \mathbb{R}^d$, $\langle a, \xi_n \rangle$ converges in distribution to $\langle a, \xi \rangle$ in $\mathbb{R}^1$.*

Lemma C.2. *Consider symmetrization of μ, i.e. $\mu^0 := \mu * \mu^-$. Then the characteristic function of μ^0 is real-valued.*

Lemma C.3. *In the case of a separable metric space,* suppμ *always exists.* $\mathrm{supp}\mu = \{\mathrm{x} \in \mathrm{X} :$ *for every open G containing x, $\mu(G) \neq 0\}$.*

Proposition C.4. *Let μ, $\nu \in \mathcal{P}(X)$. Then*

$$\mathrm{supp}(\mu * \nu) = \overline{(\mathrm{supp}\,\mu + \mathrm{supp}\,\nu)}.$$

A more general proposition is given as follows:

Proposition C.5. *Let μ be a probability on the topological space S_1 and let $f : S_1 \to S_2$ be a continuous mapping into the topological space S_2. Then*

$$f\mu = \overline{f \mathrm{supp}\,\mu}$$

In particular, for Banach spaces X and Y, probability μ on X, and a bounded linear operator $A : X \to Y$, we obtain

$$\operatorname{supp}(A\mu) = A(\operatorname{supp}\mu).$$

Proposition C.6. *Let $\mu \in \mathcal{P}(X)$. Then $(\operatorname{supp}\mu)^{\perp} = \{x^* \in X^* : \hat{\mu}(tx^*) = 1$ for all $t \in \mathbb{R}^1\}$.*

C.1 infinitely divisible and stable

Definition C.7. A probability μ on a Banach space X is said to be *infinitely divisible* if for each integer $n \geq 2$ there exists an element $\mu_n \in \mathcal{P}(X)$ such that $\mu_n^n = \mu$, where the nth power of a probability is taken in the sense of convolution.

Definition C.8. A measure $\mu \in \mathcal{P}$ is called *operator-stable* if there are a measure $\nu \in \mathcal{P}$, a sequence $\{A_n\}$ of linear operators, and a sequence $\{a_n\}$ of vectors such that

$$A_n \nu^n \delta(a_n) \Rightarrow \mu.$$

C.2 basic concepts

We say that a measure μ on $\mathbb{R}^d$ is *full* if its support is not contained in any proper hyperplane of $\mathbb{R}^d$, that is, for any x in $\mathbb{R}^d$ and any subspace W of $\mathbb{R}^d$ with $\dim W < d$, we have $\mu(W + x) < 1$.

1. The idea of fullness is the natural extension of nondegeneracy on $\mathbb{R}^1$.

2. It is shown that the set of all full measures is an open subsemigroup of $\mathcal{P}(\mathbb{R}^d)$.

Generally, the set of all full measures on $\mathbb{R}^d$ is denoted by $\mathcal{F}(\mathbb{R}^d)$. Also the set $H(\mu)$ is defined as

$$H(\mu) = \{y \in \mathbb{R}^d; \hat{\mu}(y) = 1\}.$$

Proposition C.9. *The following statements are equivalent.*

1. *μ is full.*

2. *μ^0 is full.*

3. *$H(\mu^0)$ does not contain any one-dimensional subspace.*

4. *For each $y \neq 0$, the measure $\Pi_y\mu$ is nondegenerate on $\mathbb{R}$ where $\Pi_y(x) = \langle x, y \rangle$ for $x \in \mathbb{R}^d$.*

Corollary C.10. Let $A \in \mathrm{End}(\mathbb{R}^d)$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then $A\mu$ is full if and only if A is invertible and μ is full.

For the Banach space X, let $\mathcal{A}(X)$ denote the set of all *affine* transformations on X, that is, each $\alpha \in \mathcal{A}(X)$ is given by an operator $A \in \mathrm{End(X)}$ and a vector $a \in X$, $\alpha := \langle A; a \rangle$, in the following way:

$$\alpha x := Ax + a.$$

In the same way, $r\alpha := \langle rA; ra \rangle$. The set $\mathcal{A}(X)$ is equipped by the norm

$$\|\alpha\| := \max\{\, \|A\|, \|a\| \,\},$$

then it is a Banach space.

Corollary C.11. Let $\alpha_n, \alpha \in \mathcal{A}(X)$, and assume $\alpha_n x \to \alpha x$ for all $x \in X$. Then $\mu_n \Rightarrow \mu$ in $\mathcal{P}(X)$ implies that $\alpha_n \mu_n \Rightarrow \alpha\mu$.

By $\mathcal{A}_I$, we denote the set of all invertible affine transformations on $\mathbb{R}^d$.

Corollary C.12. If $\alpha_n \mu_n \Rightarrow \mu$ with $\mu_n \in \mathcal{P}$, $\mu \in \mathcal{F}$, and $\alpha_n \in \mathcal{A}$, then $\alpha_n \in \mathcal{A}_I$ and μ_n is full for all sufficiently large n.

Next, we introduce the concept which is called "conditionally compact". (The concept is called "relatively compact" in some books.)

Definition C.13. A subset Γ of $\mathcal{P}(S)$ is called *conditionally compact* if every sequence $\{\mu_n\}$ in Γ contains a subsequence which is weakly convergent in $\mathcal{P}(S)$; the limit probability need not be in Γ.

Definition C.14. A subset Γ of $\mathcal{P}(S)$ is called *tight* if for every $\epsilon > 0$, there is a compact set K such that $\mu(K) > 1 - \epsilon$ for all $\mu \in \Gamma$.

Theorem C.15 (Prohorov Theorem)**.** *For a metric space S, every tight set Γ in $\mathcal{P}(S)$ is conditionally compact. When S is separable and complete, Γ being conditionally compact implies that Γ is tight.*

Lemma C.16. *If $\mu_n \Rightarrow \mu$ with μ full and if $\{\alpha_n \mu_n\}$ is tight, where $\alpha_n \in \mathcal{A}$, then* $\sup\|\alpha_n\| < \infty$*, that is, $\{\alpha_n\}$ is conditionally compact in $\mathcal{A}$.*

In the convergence of types theorems, a fundamental role is played by the set of operators having the property that the limit measure μ is unchanged by the action of one of these operators. More formally, we define *the invariant semigroup* of μ, $\mathrm{Inv}(\mu)$, to be

$$\mathrm{Inv}(\mu) = \{\alpha \in \mathcal{A} : \mu = \alpha\mu\}.$$

Theorem C.17. *If μ is full, then* $\mathrm{Inv}(\mu)$ *is a compact subgroup of* $\mathcal{A}_I$*. Conversely, if μ is nonfull, then* $\mathrm{Inv}(\mu)$ *is neither a group nor compact.*

Lemma C.18. *Let $\mu \in \mathcal{P}$ and $\alpha \in \mathcal{A}_I$. Then*

$$\mathrm{Inv}(\alpha\mu) = \alpha(\mathrm{Inv}(\mu))\alpha^{-1}.$$

Definition C.19. Two measures μ and ν are of the *same operator type* provided there is $\alpha \in \mathcal{A}$ such that $\mu = \alpha\nu$.

Theorem C.20. *Assume that $\beta_n\mu_n \Rightarrow \mu$, where $\beta_n \in \mathcal{A}$, $\mu_n \in \mathcal{P}$, and μ full. In order that $\alpha_n\mu_n \Rightarrow \nu$, with $\alpha_n \in \mathcal{A}$ and ν full, it is necessary and sufficient that $\nu = \alpha\mu$ for some $\alpha \in \mathcal{A}_I$, that is, μ and ν are of the same operator type, and, for all sufficiently large n,*

$$\alpha_n = \alpha\eta_n\gamma_n\beta_n,$$

where $\eta_n \to \eta_0 = \langle I; 0\rangle$ and $\gamma_n \in$ $\mathrm{Inv}(\mu)$.

C.3 Notations and assumptions

In the sequent subsection, we assume that $X, X_1, X_2, X_3, \ldots$ are i.i.d on R^d with common distribution μ and that μ belongs to the strict generalized domain of attraction of some full operator stable law ν on $\mathbb{R}^d$ with no normal component. If X belongs to the generalized domain of attraction of Y, then there exist linear operator A_n and nonrandom vectors a_n such that

$$A_n(X_1 + X_2 + \cdots + X_n) - a_n \Rightarrow Y.$$

For A_n, we say a sequence of linear operators on R^d is regularly varying with index $(-E)$ if

$$A_{[tn]}A_n^{-1} \to t^{-E},$$

for all $t > 0$. As in the Meerschaert and Scheffler (1999), the notation t^{-E} means $t^{-E} = \exp(-E\log t)$ where $\exp(A) = I + A + A^2/2! + \cdots$ is the usual exponential operator. S_n is used to be the sum of the sample,

$$S_n = X_1 + \cdots + X_n,$$

while M_n is used to represent the sample covariance matrix, i.e.

$$M_n = \sum_{i=1}^{n} X_i X_i'.$$

We give three lemmas from Meerschaert and Scheffler (1999) below.

Lemma C.21. *Suppose that μ is regularly varying with exponent E and*

$$nA_n\mu \to \phi$$

holds. If every eigenvalue of E has real part exceeding $1/2$ then

$$A_n M_n A_n^* \Rightarrow W$$

where W is infinitely divisible on M_s^d with Lévy representation $[C, 0, T\phi]$.

Lemma C.22.

$$(A_n S_n, A_n M_n A_n^*) \Rightarrow (Y, W).$$

Lemma C.23. *If $A_n S_n \Rightarrow Y$ and $A_n M_n A_n^* \Rightarrow W$ hold with $A_n = a_n^{-1} I$ then $M_n^{-1/2} S_n \Rightarrow W^{-1/2} Y$.*

If α of marginal distribution of X_1 are different, it is easy to see that we can take

$$A_n = \mathrm{d}iag\,(n^{-1/\alpha_1}, \ldots, n^{-1/\alpha_d}),$$

and E becomes

$$E = \mathrm{d}iag\,(1/\alpha_1, \ldots, 1/\alpha_d).$$

Here, we only think the case that $\alpha_i = \alpha$ for $i = 1, \ldots, d$.

We give three general lemmas, which is examined by Meerschaert and Scheffler (1999).

Lemma C.24. *Suppose that μ is regularly varying with exponent E and*

$$nA_n\mu \to \phi$$

holds. If every eigenvalue of E has real part exceeding $1/2$ then

$$A_n M_{n,Z} A_n^* \Rightarrow M$$

where M is infinitely divisible on M_s^d with Lévy representation $[C, 0, T\phi]$. Furthermore, the limit M is operator stable with exponent where $\xi M = EM + ME^*$.

Lemma C.25.

$$(A_n S_n, A_n M_{n,Z} A_n^*) \Rightarrow (Y, M).$$

Lemma C.26. *If $A_n S_n \Rightarrow Y$ and $A_n M_{n,Z} A_n^* \Rightarrow M$ hold with $A_n = a_n^{-1} I$ then $M_{n,Z}^{-1/2} S_n \Rightarrow M^{-1/2} Y$.*

Appendix D: AIC and Its Interpretation

D.1 Known Results

The approximate 1-step ahead forecast MSE is given by

$$\Sigma_{\hat{y}}(1) = \frac{T + Km + 1}{T}\tilde{\Sigma}_u, \tag{D.1}$$

where T is the number of observations, K is the dimension of vectors and m is the order of AR.

The matrix $\tilde{\Sigma}_u$ is given by

$$\tilde{\Sigma}_u = \frac{1}{T}Y(I_T - Z'(ZZ')^{-1}Z)Y' \tag{D.2}$$

from the LS estimator, which is equivalent to the ML estimator.

The LS estimator with degrees of freedom adjustment gives an unbiased estimator

$$\hat{\Sigma}_u = \frac{T}{T - Kp - 1}\tilde{\Sigma}_u. \tag{D.3}$$

D.2 FPE and AIC

The final prediction error (FPE) criterion is given by

$$\mathrm{FPE}(m) = \det\left[\,\Sigma_{\hat{y}}(1)\,\right] = \det\left[\frac{T + Km + 1}{T}\frac{T}{T - Km - 1}\tilde{\Sigma}_u(m)\right] \tag{D.4}$$

A very similar criterion abbreviated by AIC is defined as

$$\mathrm{AIC(m)} = \ln|\tilde{\Sigma}_u(m)| + \frac{2}{T}(\mathrm{numberoffreelyestimatedparameters}) \tag{D.5}$$

The connection is understood by

$$\ln\mathrm{FPE}(m) = \mathrm{AIC(m)} + \frac{2m}{T} + O(T^{-2}), \tag{D.6}$$

but the constant term must be suspected since the choice of the number of freely estimated parameters.

D.3 Generalized Information Criterion

First, we define the tool in the argument:

- A class of parametric models

$$\{f_\theta;\ \dim\theta = p, p = 0, 1, \dots, L\}; \tag{D.7}$$

- A fundamental parametric model f_{θ_0} with order $\dim\theta_0 = p_0$;
- The true model g, which is contiguous to the fundamental parametric model

$$g = f_{(\theta_0, h/\sqrt{n})}, \quad \text{where } h = (h_1, \dots, h_{K-p_0})'. \tag{D.8}$$

To well define f_{θ_0}, we use a measure of disparity D and θ_0 is defined as

$$D(f_{\theta_0}, g) = \min_{\theta\in\Theta} D(f_\theta, g). \tag{D.9}$$

Then GAIC is defined as

$$\mathrm{G}AIC(p) = nD(f_{\hat{\theta}_n}, \hat{g}_n) + p, \tag{D.10}$$

which is possible to use in i.i.d case, ARMA case, regression case, or CHARN case.

Before the result, we need some assumptions:

(D1) $f_\theta(\cdot)$ is continuously twice differentiable with respect to θ.

(D2) The fitted model is nested, i.e., $\theta(p+1) = (\theta(p)', \theta_{p+1})'$.

(D3) As $n \to \infty$, the estimator $\hat{\theta}_n$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}), \tag{D.11}$$

and $I(\theta) = J(\theta)$ if $g = f_\theta$.

(D4) If $g = f_{\theta(p_0)}$, then $D(f_{\theta(p)}, g)$ is uniquely minimized at $p = p_0$.

Define $\alpha_i = P(\chi_i^2 > 2i)$ and

$$\rho(l, \{\alpha_i\}) = \sum_l^{*} \Big\{ \prod \Big\}_{i=1}^{l} \frac{1}{r!} \Big(\frac{\alpha_i}{i}\Big)^{r_i} \}, \tag{D.12}$$

where Σ_l^* extends over all l-tuples $(r_1, \dots, r_l)$ of non-negative integers satisfying $r_1 + 2r_2 + \cdots + lr_l = l$.

Theorem D.1. *Suppose $g = f_\theta(p_0)$. Then, under Assuption (D1)-(D4), the asymptotic selection probability of $\hat{p}$ selected by $GAIC(p)$ is given by*

$$\lim_{n\to\infty} P(\hat{p} = p) = \begin{cases} 0, & 0 \le p < p_0, \\ \rho(p - p_0, \{\alpha_i\})\rho(L - p, \{1 - \alpha_i\}), & p_0 \le p \le L \end{cases} \tag{D.13}$$

where $\rho(0, \{\alpha_i\})\rho(0, \{1 - \alpha_i\}) = 1$.

To see the asymptotic selection probability under the true model, we need another set of assumptions:

(D5) If $h = 0$, the model $f_\theta(\cdot)$ satisfies (D1) - (D4).

(D6) $\{P^{(n)}_{\theta_0,h/\sqrt{n}}\}$ has the LAN property, i.e.

$$\log \frac{dP^{(n)}_{(\theta_0,h/\sqrt{n})}}{dP^{(n)}_{(\theta_0,0)}} = h'\Delta_n(\theta_0) - \frac{1}{2}h'\Gamma(\theta_0)h + o_{p_0}(1), \tag{D.14}$$

$$\Delta_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Gamma(\theta)_0), \quad \text{under } P^{(n)}_{(\theta_0,0)}. \tag{D.15}$$

(D7) Under $P^{(n)}_{(\theta_0,0)}$, for $p_0 < l \le L$, the random vectors $(Z_1^{(l)'}, \Delta_n')'$ are asymptotically normal with

$$\lim_{n\to\infty} \mathrm{Cov}(Z_1^{(l)}, \Delta_n) = \{\sigma_{jk}(l);\ j = 1,\dots,l, k = 1,\dots,K - p_0\}, \tag{D.16}$$

and

$$\lim_{n\to\infty} \mathrm{Var}(Z_1^l) = \begin{pmatrix} F_{11}^{(l-1)} & F_{12}^{(l-1)} \\ F_{21}^{(l-1)} & F_{22}^{(l)} \end{pmatrix}. \tag{D.17}$$

Write

$$\Sigma^{(l-1)} = \begin{pmatrix} \sigma_{11}(l) & \dots & \sigma_{1,K-p_0}(l) \\ \vdots & & \vdots \\ \sigma_{l-1,1}(l) & \dots & \sigma_{l-1,K-p_0}(l) \end{pmatrix}, \tag{D.18}$$

and $\sigma^{(l)} = (\sigma_{l,1}(l), \dots, \sigma_{l,K-p_0}(l))'$. Then for

$$W_1, \dots, W_{L-p_0} \sim \text{i.i.d } \mathcal{N}(0,1), \tag{D.19}$$

we define

$$\begin{aligned}
\mu_l &= (\sigma^{(l)'}h - F_{21}^{(l-1)}F_{11}^{(l-1)^{-1}}\Sigma^{(l-1)}h)\sqrt{F_{22\cdot 1}^l}, \quad l = p_0+1,\dots,L, \\
S_{j,m} &= (W_m - \mu_{m+p_0})^2 + \dots + (W_{m-j+1} - \mu_{m+p_0-j+1})^2 - 2j, \\
&\qquad 1 \le m \le L - p_0,\ j = 1,\dots,m, \\
T_{j,k} &= (W_{k+1} - \mu_{k+p_0+1})^2 + \dots + (W_{k+j} - \mu_{k+p_0+j})^2 - 2j, \\
&\qquad 0 \le k \le L - p_0 - 1,\ j = 1,\dots,L - p_0 - k.
\end{aligned}$$

Then we have the following theorem.

Theorem D.2. *Assume Assumption (G1)-(G7). Under $P^{(n)}_{(\theta_0,h/\sqrt{n})}$, the asymptotic selection probability of $\hat{p}$ is given by*

$$\lim_{n\to\infty} P^{(n)}_{\theta_0,h/\sqrt{n}}\{\hat{p}=p\} = \begin{cases} 0, & 0 \le p < p_0 \\ \beta(p-p_0)\gamma(p-p_0) & p_0 \le p \le L, \end{cases} \tag{D.20}$$

where

$$\beta(m) = P\left\{\bigcap_{j=1}^{m}(S_{j,m}>0)\right\}, \quad \gamma(k) = P\left\{\bigcap_{j=1}^{L-p_0-k}(T_{j,k}\le 0)\right\}, \tag{D.21}$$

$$\beta(0)=\gamma(0)=1. \tag{D.22}$$

Appendix E: LAN, LAM and LAMN

The theory of LAN started from the idea of Le Cam, who was the major figure in the development of abstract general asymptotic theory in mathematical statistics. First, we review his work to express our respect for him.

E.1 Le Cam's lemmas

The likelihood ratio

Introduce the likelihood ratio $L_n = q_n/p_n$, or more precisely,

$$L_n(\boldsymbol{x}) = \begin{cases} \dfrac{q_n(\boldsymbol{x})}{p_n(\boldsymbol{x})} & \text{if } p_n(\boldsymbol{x}) > 0 \\ 1 & \text{if } q_n(\boldsymbol{x}) = p_n(\boldsymbol{x}) = 0 \\ \infty & \text{if } q_n(\boldsymbol{x}) > p_n(\boldsymbol{x}) = 0. \end{cases}$$

The most remarkable three lemmas are in the flow of [contiguity → LAN → Alternative Hypotheses]. Let F_n be the distribution function of L_n under P^n:

$$F_n(x) = P^n(L_n \le x).$$

Lemma E.1. *Assume that F_n converges weakly (at continuity points) to a distribution function F such that*

$$\int_0^\infty x\,dF(x) = 1.$$

Then the densities q_n are contiguous to the densities p_n, $n \ge 1$.

Proof. **Step 1** For any test statistics Ψ_n, we decompose it like

$$\begin{aligned}\int \Psi_n \, dQ_n &= \int_{\{L_n \le y\}} \Psi_n \, dQ_n + \int_{\{L_n > y\}} \Psi_n \, dQ_n \\ &\le y \int \Psi_n \, dP_n + 1 - \int_0^y x \, dF_n.\end{aligned}$$

Step 2 $\forall \epsilon > 0, \exists y > 0$ (a continuity point) such that

$$1 - \int_0^y x \, dF < \frac{1}{2}\epsilon.$$

Step 3 $F_n \to F$ at continuity point entails

$$\int_0^y x \, dF_n \to \int_0^y x \, dF.$$

Step 4

$$1 - \int_0^y x \, dF_n < \frac{1}{2}\epsilon \quad \text{for } n \ge n_0.$$

Step 5 $\int \Psi_n \, dP_n \to 0$ entails

$$y \int \Psi_n \, dP_n < \frac{1}{2}\epsilon, \quad \text{for } n \ge n_1$$

Step 6 From (4) and (5), it is easily seen that

$$\int \Psi_n \, dQ_n \to 0.$$

□

Corollary E.2. If, under P_n, the ratio L_n is asymptotically log-normal $(-\frac{1}{2}\sigma^2, \sigma^2)$, then the densities q_n are contiguous to the densities p_n.

Proof. Since Y is distributed as log-normal (μ, σ^2), $\log Y$ is distributed as normal distribution.

$$EY = E \exp(\log Y) = \exp(\mu + \frac{1}{2}\sigma^2).$$

As what we see at the lemma above, if $\mu = -\frac{1}{2}\sigma^2$, then q_n are contiguous to the densities p_n. □

Lemma E.3. *Suppose that the following condition (the UAN (uniform asymptotic negligibility) condition) holds:*

$$\max_{1\le i\le n} P^n(\left|\frac{g_{ni}}{f_{ni}}-1\right| > \epsilon) \to 0.$$

Let W_n *be*

$$W_n = 2\sum_{i=1}^{N_n}\{[\, g_{ni}(X)/f_{ni}(X_i)\,]^{\frac{1}{2}} - 1\}.$$

Suppose also that W_n *converges in distribution to* $N(-\sigma^2/4, \sigma^2)$ *for some* σ^2*. Then,*

$$\log L_n - (W_n - \sigma^2/4) = o_{P^n}(1)$$

and hence

$$\log L_n \xrightarrow{d} \mathcal{N}(-\frac{\sigma^2}{2}, \sigma^2).$$

Proof. Note that

$$h(x) = h(x_0) + (x-x_0)h'(x_0) + \frac{1}{2}(x-x_0)^2\int_0^1 2(1-\lambda)h''[\,x_0+\lambda(x-x_0)\,]d\lambda.$$

Let T_{ni} satisfies

$$T_{ni} = 2[\, g_{ni}(X_i)/f_{ni}(X_i)\,]^{\frac{1}{2}} - 2.$$

Step 1

$$\log\frac{g_{ni}}{f_{ni}} = T_{ni} - \frac{1}{4}T_{ni}^2\int_0^1 [\,[2(1-\lambda)/(1+\frac{1}{2}\lambda T_{ni})^2\,]d\lambda.$$

Step 2

$$\log L_n = W_n - \frac{1}{4}\sum_{i=1}^{N_n} T_{ni}^2\int_0^1 [\,[2(1-\lambda)/(1+\frac{1}{2}\lambda T_{ni})^2\,]d\lambda.$$

Step 3 Introduce

$$T_{ni}^{\delta} = \begin{cases} T_{ni}, & \text{if } |T_{ni}| \le \delta, \\ 0, & \text{otherwise.}\end{cases}$$

Then we can have a lemma from (Loève (1955)).

Lemma E.4. *If W_n is asymptotically distributed as normal$(-\frac{1}{4}\sigma^2, \sigma^2)$. Then for every $\delta > 0$,*

$$\begin{aligned} \sum_{i=1}^{N_n} P^n(|T_{ni}| > \delta) &\to 0, \\ \sum_{i=1}^{N_n} ET_{ni}^{\delta} &\to -\frac{1}{4}\sigma^2, \\ \sum_{i=1}^{N_n} \mathrm{Var} T_{ni}^{\delta} &\to \sigma^2. \end{aligned}$$

The Lemma will give a good picture of the proof for the Lemma.

□

Lemma E.5. *Assume that the pair $(S_n, \log L_n)$ is under P_n asymptotically jointly normal $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ with $\mu_2 = \frac{1}{2}\sigma_2^2$. Then, under Q_n, S_n is asymptotically normal $(\mu_1 + \sigma_{12}, \sigma_1^2)$.*

Proof. **Step 1** Write

$$Q_n(S_n \le x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} e^n \, dF_n(u, v) + Q_n(p_n = 0, S_n \le x),$$

where $F_n(u, v)$ denotes the distribution function of $(S_n \log L_n)$.

Step 2 $\mu_2 = -\frac{1}{2}\sigma_2^2$ implies the contiguity, and hence

$$Q_n(p_n = 0, S_n \le x) \to 0,$$

since $P_n(p_n = 0, S_n \le x) = 0 \to 0$.

Step 3

$$\int_{-\infty}^{x} \int_{-c}^{c} e^v \, dF_n(u, v) \to \int_{-\infty}^{x} \int_{-c}^{c} e^v d\Phi(u, v).$$

Step 4 From Step 1, 2 and 3,

$$Q_n(S_n \le x) \to \int_{-\infty}^{x} \int_{-\infty}^{\infty} e^v \, d\Phi(u, v),$$

which implies the conclusion.

□

E.2 Extension of Le Cam's third lemma

Theorem E.6 (Van der Vaart Theorem 6.6)**.** *Let X_n be a map as $X_n : \Omega \to \mathbb{R}^k$. Then if Q_n is contiguous to P_n, and $(X_n, \frac{dQ_n}{dP_n}) \xrightarrow{d} (X, V)$ under P_n, then we have $X_n \xrightarrow{d} L$ under Q_n where*

$$L(B) = E1_B(X)V,$$

and L defines a probability measure.

Suppose $\alpha \in [1, 2)$ and

$$\begin{aligned} X &\sim S(\alpha, 0, \Sigma, \delta); \quad S(\cdot) \text{ denotes a stable distribution,} \\ V &\sim \log\mathcal{N}(-\frac{1}{2\sigma^2}, \sigma^2); \quad \log\mathcal{N} \text{ denotes a lognormal distribution.} \end{aligned}$$

Theorem E.7 (The main result)**.** *Let X_n be a map as $X_n : \Omega \to \mathbb{R}^k$. Then if Q_n is contiguous to P_n, and $(X_n, \frac{dQ_n}{dP_n}) \xrightarrow{d} (X, V)$ under P_n, where X is a stable random vector of exponent α, and V is also defined as above. then we have $X_n \xrightarrow{d} L$ under Q_n where L is also a stable random vector of exponent α.*

Proof. Let $L(B) = E1_B(X)e^W$, where W is a normal random variable, since V is lognormal. This shows the logarithm of the likelihood ratio is as normal distributed and for this reason, it is easy to see that Q_n is contiguous to P_n.

$$\int e^{iu'x} dL(x) = Ee^{iu'X}e^W.$$

Since (X, W) is under a disjoint distribution of stable and normal as expectation above, to calculate the right hand side in the equation above, we just only have to let

$$t' = (u', -i),$$

then

$$\begin{aligned} Ee^{iu'X}e^W &= e^{iu'\mu - \frac{1}{2}\sigma^2 - \frac{1}{2}(u',-i)\Sigma^*(u',-i)'} \\ &= \exp\{iu'(\delta + \tau) - \frac{1}{2}u'\Sigma u\} \end{aligned}$$

Since the right hand side shows the characteristic function of multivariate stable distribution (formally, it is called Elliptically contoured multivariate stable distribution $S(\alpha, 0, \Sigma, \delta + \tau)$), we have the conclusion. ∎

E.3 Introduction

Let $X_0, X_1, \dots, X_n$ be i.i.d random vectors defined on the probability space $(\mathcal{X}, \mathcal{A}, P_\theta)$, and $\theta \in \Theta$ open $\subset \mathbb{R}^k$, $k \geq 1$. For the observations, let $\mathcal{A}_n = \sigma(X_0, X_1, \dots, X_n)$ be the σ-field induced by them, and let $P_{n,\theta} = P_\theta \,|\, \mathcal{A}_n$ be the restriction of P_θ to $\mathcal{A}_n$. We set

$$q(X_0; \theta, \theta^*) = \frac{dP_{0,\theta^*}}{dP_{0,\theta}},$$

be the Radon-Nikodym derivative involved and

$$\varphi_j(\theta, \theta^*) = \varphi(X_j, \theta, \theta^*) = [\, q(X_j; \theta, \theta^*)\,]^{1/2},$$

so that $\varphi_j(\theta, \theta^*)$ is square $P_{0,\theta}$-integrable.

Let a sequence $\{\theta_n\}$, which is close to θ, be defined by

$$\theta_n = \theta + h_n/\sqrt{n}, \quad \text{with } h_n \to h \in \mathbb{R}^d. \tag{E.1}$$

Remark E.8. Usually, $P_{0,\theta} = \mu$.

To see the assumptions for LAN, we need first look at two definitions here.

Definition E.9 (q.m.d)**.** The family $\{P_\theta, \theta \in \Theta\}$ is quadratic mean differentiable(q.m.d) at θ_0 if there exists a vector of real-valued functions $\eta(\cdot, \theta_0) = (\eta_1(\cdot, \theta_0, \dots, \eta_k(\cdot, \theta_0))')$ such that

$$\int_{\mathcal{X}} \left[\sqrt{p_{\theta_0+h}(x)} - \sqrt{p_{\theta_0}(x)} - < \eta(x, \theta_0), h > \right]^2 d\mu(x) = o(|h|^2)$$

as $|h| \to 0$.

Definition E.10 (Fisher Imformation matrix)**.** For a q.m.d family with derivative $\eta(\cdot, \theta)$, define the Fisher information matrix to be the matrix $\Gamma(\theta_0)$ with (i, j) entry

$$\Gamma_{i,j}(\theta_0) = 4 \int \eta_i(x, \theta_0)\eta_j(x, \theta_0)\, d\mu(x).$$

Remark E.11. $\eta(x, \theta_0)$ can be seen as

$$\eta(x, \theta_0) = \frac{1}{2}\dot{l}_\theta \sqrt{p_\theta},$$

where

$$\dot{l}_\theta(x) = 2\frac{1}{\sqrt{p_\theta(x)}}\frac{\partial}{\partial\theta}\sqrt{p_\theta(x)} = \frac{\partial}{\partial\theta}\log p_\theta(x).$$

We summarize this point in the following theorem.

Theorem E.12. *Suppose Θ is an open subset of $\mathbb{R}$ and fix $\theta_0 \in \Theta$. Assume $p_\theta^{1/2}(x)$ is an absolutely continuous function of θ in some neighborhood of θ_0, for μ-almost all x. Also, assume for μ-almost all x, the derivative $p'_\theta(x)$ of $p_\theta(x)$ with respect to θ exists at $\theta = \theta_0$. Define*

$$\eta(x,\theta) = \begin{cases} \dfrac{p'_\theta(x)}{2p_\theta^{1/2}(x)} & \text{if } p_\theta(x) > 0 \text{ and } p'_\theta(x) \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

Also, assume the Fisher Information $\Gamma(\theta)$ is finite and continuous in θ at θ_0. Then, $\{P_\theta\}$ is q.m.d at θ_0 with quadratic mean derivative $\eta(\cdot,\theta_0)$.

Proof. **Step 1** Let x be such that $p_\theta^{1/2}(x)$ is absolutely continuous in $[\theta_0, \theta_0+\delta]$, then

$$\left\{\frac{1}{\delta}[\,p_{\theta_0+\delta}^{1/2}(x) - p_{\theta_0}^{1/2}(x)\,]\right\} = \frac{1}{\delta^2}\left[\int_0^\delta \eta(x,\theta_0+\lambda)\,d\lambda\right]^2 \le \frac{1}{\delta}\int_0^\delta \eta^2(x,\theta_0+\lambda)\,d\lambda.$$

Step 2 From Step 1,

$$\int \left\{\frac{1}{\delta}[\,p_{\theta_0+\delta}^{1/2}(x) - p_{\theta_0}^{1/2}(x)\,]\right\} d\mu(x) \le \frac{1}{4\delta}\int_0^\delta I(\theta_0+\lambda)d\lambda.$$

Step 3 By continuity of $I(\theta)$ at θ_0, the right hand side tends to $\frac{1}{4}I(\theta_0)$ as $\delta \to 0$.

Step 4 From the definition of η, we can see that

$$\frac{1}{\delta}[\,p_{\theta_0+\delta}^{1/2}(x) - p_{\theta_0}^{1/2}(x)\,] \to \eta(x,\theta_0).$$

Step 5 From Vitali's Theorem,

$$\int_{\mathcal{X}} \left[\sqrt{p_{\theta_0+\delta}(x)} - \sqrt{p_{\theta_0}(x)} - <\eta(x,\theta_0),\delta>\right]^2 d\mu(x) \to 0$$

as $\delta \to 0$.

□

Definition E.13. If the Fisher information matrix is nonsingular at $\theta_0 \in \Theta^\circ$, then we call θ_0 a *regular point.* Furthermore, if every point of Θ is regular and $\eta_i(\cdot)$ is continuous, then we call the parametrization $\theta \to P_\theta$ is regular and $\boldsymbol{P} = \{P_\theta\}$ a regular parametric model.

For simplicity, we will eliminate x in the notation and focus on θ instead. The other notations in this section are given below:

$$\begin{aligned}
L_n(\theta,\theta_n) &= \frac{dP_{n,\theta_n}}{dP_\theta} \quad \text{(the relevant likelihood ratio);} \\
\Lambda_n(\theta,\theta^*) &= \log L_n(\theta,\theta^*) \quad \text{(log-likelihood ratio);} \\
\Delta_n(\theta) &= \frac{2}{\sqrt{n}}\sum_{j=0}^{n}\dot{\varphi}_j(\theta); \quad \text{(score function)} \\
\Gamma(\theta) &= 4E_\theta\left[\,\dot{\varphi}_0(\theta)\dot{\varphi}_0'(\theta)\,\right]; \quad \text{(Fisher information)} \\
A(h,\theta) &= \frac{1}{2}h'\Gamma(\theta)h.
\end{aligned}$$

E.4 Assumtions (A1)-(A4)

To obtain LAN, we need assumptions as follows:

(A1) The probability measures $\{P_{0,\theta};\, \theta \in \Theta\}$ are mutually absolutely continuous.

(A2) (a) For each $\theta \in \Theta$, the random function $\varphi_0(\theta,\theta^*)$ is differentiable in quadratic mean with respect to θ^* at θ when P_θ is employed.

(b) $\dot{\varphi}_0(\theta)$ is $\mathcal{A}_0 \times \mathcal{C}$-measurable, where $\mathcal{C}$ is the σ-field of Borel subsets of Θ.

(A3) $\Gamma(\theta)$ is positive definite for every $\theta \in \Theta$.

(A4) For each $\theta \in \Theta$,

(a) $q(X_0;\theta,\theta^*) \to 1$ in $P_{0,\theta}$-probability, as $\theta^* \to \theta$.

(b) $q(X_0;\theta,\theta^*)$ is $\mathcal{A}_0 \times \mathcal{C}$-measurable.

Theorem E.14. *Let θ_n, $L_n(\theta,\theta_n)$, $\Delta_n(\theta)$, and $A(h,\theta)$ be defined above. Then, under assumptions* (E.4)*,*

$$\Lambda_n(\theta,\theta_n) - h'\Delta_n(\theta) \xrightarrow{P_{n,\theta}} -A(h,\theta); \tag{E.2}$$

$$\mathcal{L}\left[\,\Delta_n(\theta)\,|\,P_{n,\theta}\,\right] \Rightarrow \mathcal{N}(0,\Gamma(\theta)); \tag{E.3}$$

$$\mathcal{L}\left[\,\Lambda_n(\theta,\theta_n)\,|\,P_{n,\theta}\,\right] \Rightarrow \mathcal{N}(-\frac{1}{2}h'\Gamma(\theta)h, h'\Gamma(\theta)h). \tag{E.4}$$

Remark E.15. Theorem (E.14) shows the relationship between log-likelihood ratio, central sequence and Fisher information under the true distribution. Central sequence follows CLT to have asymptotic normality and so if Fisher information is non-random, then the log-likelihood ratio has also asymptotic normality.

Remark E.16. Every p_θ in a regular parametric model $\boldsymbol{P}$ satisfy the assumptions above.

Theorem E.17 (ULAN)**.** *Suppose that* $\boldsymbol{P} = \{P_\theta : \theta \in \Theta\}$ *is a regular parametric model, then we have*

$$\Lambda_n(\theta, \theta_n) = h'\Delta_n(\theta) - \frac{1}{2}h'\Gamma(\theta)h + R_n(\theta, h),$$

where $R_n(\theta, h) \to o_{P_\theta}(1)$ *uniformly for* $\theta \in K$ *compact* $\subset \Theta$ *and* $|h| \leq M$*.*

On the contrary, the distribution of log-likelihood ratio under the alternative hypotheses can be also obtained through the proposition below:

Proposition E.18. *Let* $\{h_n^*\}$*be a bounded sequence in* $\mathbb{R}^k$*, and set* $\theta_n^* = \theta + h_n^*/\sqrt{n}$*. Then under assumptions* (E.4)*, the sequence* $\{P_{n,\theta}\}$ *and* $\{P_{n,\theta_n^*}\}$ *are contiguous; in particular, so are the sequences* $\{P_{n,\theta}\}$ *and* $\{P_{n,\theta_n}\}$*.*

Theorem E.19. *Under assumptions* (E.4)*,*

$$\Gamma_n(\theta, \theta_n) - h'\Delta_n(\theta) \xrightarrow{P_{n,\theta_n}} -A(h, \theta); \tag{E.5}$$

$$\mathcal{L}\left[\,\Lambda_n(\theta, \theta_n) \,|\, P_{n,\theta_n}\,\right] \Rightarrow \mathcal{N}(\frac{1}{2}h'\Gamma(\theta)h, h'\Gamma(\theta)h); \tag{E.6}$$

$$\mathcal{L}\left[\,\Delta_n(\theta) \,|\, P_{n,\theta_n}\,\right] \Rightarrow \mathcal{N}(\Gamma(\theta)h, \Gamma(\theta)). \tag{E.7}$$

Remark E.20. Another version for the result is

$$\Delta_n(\theta_n) - \Delta_n(\theta) \to_{p_\theta} -\Gamma(\theta)h, \quad n \to \infty, \tag{E.8}$$

uniformly in $\theta \in K$ and $|t| \leq M$. In other words, the difference between two scores are asymptotically linear where the slope is the Fisher information.

The likelihood ratio behaves as if it were an exponential family.

Theorem E.21. *Under assumptions* (E.4)*, there exists a truncated version* $\Delta_n^*(\theta)$ *of* $\Delta_n(\theta)$ *such that:*

$$E_\theta \mathbb{E}^{h'\Delta_n^*(\theta)} \equiv \mathbb{E}^{B_n(h)} < \infty, \tag{E.9}$$

$$P_{n,\theta}\left[\,\Delta_n^*(\theta) \neq \Delta_n(\theta)\,\right] \to 0, \tag{E.10}$$

$$P_{n,\theta_n}\left[\,\Delta_n^*(\theta) \neq \Delta_n(\theta)\,\right] \to 0, \tag{E.11}$$

and if

$$R_{n,h}(A) = \mathbb{E}^{-B_n(h)} \int_A \mathbb{E}^{h'\Delta_n^*(\theta)} dP_{n,\theta}, \quad A \in \mathcal{A}_n,$$

then

$$\|P_{n,\theta_n} - R_{n,h_n}\| \to 0, \tag{E.12}$$

or

$$\sup\left\{\,\|P_{n,\theta_n} - R_{n,h}\|;\ h \in B \textit{ bounded} \subset \mathbb{R}^k\,\right\} \to 0. \tag{E.13}$$

Remark E.22. The truncated version $\Delta_n^*(\theta)$, which can be seen as exponential family, can not be identified from $\Delta_n(\theta)$ under either hypothesis or alternatives.

E.5 LAN for stochastic Process

For stochastic, we refer to Taniguchi and Kakizawa [2000].

Theorem E.23 (Le Cam (1986)). *Suppose that under $P_{0,n}$ the following conditions (L1)-(L4) are satisfied.*

(L1) $\max_k |Y_{n,k}| \to_p 0$,

(L2) $\sum_k Y_{n,k}^2 \to_p \sigma^2/4$,

(L3) $\sum_k E(Y_{n,k}^2 + 2Y_{n,k}|\mathcal{F}_{n,k-1}) \to_p 0$,

(L4) $\sum_k E\,(Y_{n,k}^2 1_{Y_{n,k}>\delta|\mathcal{F}_{n,k-1}}) \to_p 0$ *for some* $\delta > 0$. *Then* $\Lambda_n \to_d \mathcal{N}(-\sigma^2/2, \sigma^2)$.

Swensen (1985) gave a martingale difference version to Theorem E.23, and apply the new version to the AR model. Kreiss (1987, 1990) generalized the result to the ARMA process.

In Taniguchi and Kakizawa [2000], they posed the following assumption to the m-vector linear process

$$\boldsymbol{X}(t) = \sum_{j=0}^{\infty} A_\theta(j)\boldsymbol{U}(t-j), \quad t \in \boldsymbol{Z}, \tag{E.14}$$

where the $\boldsymbol{U}(t)$ are i.i.d. m-vector random variables with probability density $p(\boldsymbol{u}) > 0$ on $\mathbb{R}^m$, and $A_\theta(j)$, $j \in \mathbb{Z}$, are matrices depending on a parameter vector $\theta \in \Theta$ compact $\subset \mathbb{R}^q$.

(TK 1) Suppose$\det A_\theta(z) \neq 0$ for $|z| \leq 1$ and $A_\theta(z)^{-1}$ can be expanded by

$$A_\theta(z)^{-1} = I_m + \sum_{j=1}^{\infty} (B_\theta(j) z^j)^j. \tag{E.15}$$

(a) For some D $(0 < D < 1/2)$, the coefficient matrices $A_\theta(j)$ satisfy

$$|A_\theta(j)_{kl}| = O(j^{-1+D}), \quad j \in \mathbb{N}, \quad \text{for all } 1 \leq k, l \leq m. \tag{E.16}$$

(b) Every $A_\theta(j)$ is continuously twice differentiable with respect to θ, and the derivatives satisfy

$$|\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} A_\theta(j)_{ab}| = O(j^{-1+D}(\log j)^k), \quad k = 0, 1, 2 \tag{E.17}$$

for $a, b = 1, \ldots, m$, where $\partial_i = \partial/\partial\theta_i$.

(c) where

$$|B_\theta(j)_{kl}| = O(j^{-1-D}), \quad j \in \mathbb{N}, \quad \text{for all } 1 \leq k, l \leq m. \tag{E.18}$$

(d) Every $B_\theta(j)$ is continuously twice differentiable with respect to θ, and the derivatives satisfy

$$|\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k} B_\theta(j)_{ab}| = O(j^{-1-D}(\log j)^k), \quad k = 0, 1, 2 \tag{E.19}$$

for $a, b = 1, \dots, m$.

(TK 2) $p(\cdot)$ satisfies

$$\lim_{\|\boldsymbol{u}\|\to\infty} p(\boldsymbol{u}) = 0, \quad E(\boldsymbol{u}) = 0, \quad \text{Var}(\boldsymbol{u}) = I_m. \tag{E.20}$$

(TK 3) The continuous derivative Dp of $p(\cdot)$ exists on $\mathbb{R}^m$.

(TK 4) $\int_{\mathbb{R}^m} |\phi(\boldsymbol{u})_{kl}|^4 p(\boldsymbol{u}) d\boldsymbol{u} < \infty$, for all $1 \leq k, l \leq m$, where $\phi(\boldsymbol{u}) = p^{-1}Dp$.

Theorem E.24. *Suppose that assumptions above hold. Write*

$$B_{h'\partial\theta}(j) = \sum_{t=1}^{q} h_l \partial_l B_\theta(j), \tag{E.21}$$

$$\mathcal{F}(p) = \int_{\mathbb{R}^m} \phi(\boldsymbol{u})\phi(\boldsymbol{u})' p(\boldsymbol{u}) d\boldsymbol{u}. \tag{E.22}$$

Then

$$\Lambda_n(\theta, \theta_n) = \Delta_{n,h}(\theta) - \frac{1}{2}\Gamma_h(\theta) + o_{p_\theta}(1), \tag{E.23}$$

where

$$\Delta_{n,h}(\theta) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \phi(\boldsymbol{U}(k))' \sum_{j=1}^{k-1} B_{h'\partial\theta}(j) \boldsymbol{X}(k-j), \tag{E.24}$$

$$\Gamma_h(\theta) = \text{tr}\{\mathcal{F}(p) \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} B_{h'\partial\theta}(j_1) R(j_1 - j_2) B_{h'\partial\theta}(j_2)'\}. \tag{E.25}$$

Further, under P_θ, we have

$$\Delta_{n,h}(\theta) \xrightarrow{P_{n,\theta}} \mathcal{N}(0, \Gamma_h(\theta)). \tag{E.26}$$

Finally, For all $n \in \mathbb{N}$, and all $h \subset \mathcal{H} \subset \mathbb{R}^q$, the mapping $h \to P_{n,\theta_n}$ is continuous with respect to the variational distance

$$\|P - Q\| = \sup\{|P(A) - Q(A)|| : A \in \mathcal{A}_n\}. \tag{E.27}$$

E.6 Metrics for the space of measure

Definition E.25. A set $\mathcal{P}$ is a metric space if there exists a real-valued function d defined on $\mathcal{P} \times \mathcal{P}$ such that for all points p, q and r in $\mathcal{P}$,

- $d(p,q) \geq 0$;
- $d(p,q) = d(q,p)$;
- $d(p,q) \leq d(p,r) + d(r,q)$.

A function d satisfying these conditions is called a metric.

Example 2. The Lévy distance:

$$\rho_L(F,G) \equiv \inf\{\epsilon > 0; F(x-\epsilon) \leq G(x) \leq F(x+\epsilon) + \epsilon \quad \textit{for all } x\}.$$

note.

$$F_n \xrightarrow{d} F \iff \rho_L(F_n, F) \to 0.$$

Example 3 (Uniform metric)**.**

$$d_K(F,G) \equiv \sup_t |F(t) - G(t)|.$$

note. Let $\hat{F}_n$ be the empirical c.d.f. defined by

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \leq t\}.$$

Consider the problem of testing the simple null hypothesis that $F = F_0$ versus $F \neq F_0$. The Kolmogorov-Smirnov goodness of fit test statistic is given by

$$T_n \equiv \sup_{t \in \mathbb{R}} n^{1/2} |\hat{F}_n(t) - F_0(t)| = n^{1/2} d_K(\hat{F}_n, F_0).$$

The statistic tends to small under the null bypothesis and large under the alternative, which is justified by the theorem below. The Kolmogorov-Smirnov test rejects the null hypothesis if $T_n > s_{n,1-\alpha}$, where $s_{n,1-\alpha}$ is the $1-\alpha$ quantile of the null distribution of T_n when F_0 is the uniform $U(0,1)$ distribtuion.

Theorem E.26 (Dvoretzky, Kiefer, Wolfowitz Inequality)**.** *Suppose $X_1, \ldots, X_n$ are i.i.d. real-valued random variables with c.d.f. F. Let $\hat{F}_n$ be the empirical c.d.f. Then, for any $d > 0$ and any positive integer n,*

$$P\{d_K(\hat{F}_n, F) > d\} \leq C \exp(-2nd^2),$$

where C is a universal constant.

In fact, the result is improved by Massart(1990) as

$$P(\sqrt{n}\sup_x |\hat{F}_n(x) - F(x)| > \lambda) \le 2\exp(-2\lambda^2),$$

which has not any restriction on λ.
note. There are so many technical calculations and lemmas in the paper, so we will not get involved in the proof.

Using the fact, it is easy to show the theorem below.

Theorem E.27 (Glivenko-Cantelli Theorem). *Suppose $X_1, \ldots, X_n$ are i.i.d. real-valued random variable with c.d.f. F. Then*

$$d_K(\hat{F}_n, F) \to 0 \quad a.s..$$

E.7 L_1 norm, L_2 norm and contiguity

For a test ϕ, denote the sum of the probability of rejecting P_0 when P_0 is true and the probability of rejecting P_1 when P_1 is true by

$$S_{P_0,P_1}(\phi) = \int_{\mathcal{X}} \phi(x)\, dP_0(x) + \int_{\S} (1 - \phi(x))\, dP_1(x)$$

and let

$$S(P_0, P_1) = \inf_\phi [\, S_{P_0,P_1}(\phi)\,].$$

Definition E.28. The total variation distance between P_0 and P_1, denoted $||P_1 - P_0||_1$, is given by

$$||P_1 - P_0||_1 = \int |p_1 - p_0|\, d\mu,$$

where p_i is the density of P_i with respect to any measure μ dominating both P_0 and P_1.

Definition E.29. Let P_0 and P_1 be probabilities on $(\mathcal{X}, \mathcal{F})$. The Hellinger distance $H(P_0, P_1)$ between P_0 and P_1 is given by

$$H^2(P_0, P_1) = \frac{1}{2}\int_{\mathcal{X}} \left[\sqrt{p_1(x)} - \sqrt{p_0(x)}\right]^2 d\mu(x).$$

Furthermore, define

$$\rho(P_0, P_1) = 1 - H^2(P_0, P_1).$$

then,

$$\rho(P_0, P_1) = \int_{\mathcal{X}} \sqrt{p_0(x)p_1(x)}\, d\mu(x).$$

note. Let P^n denote the joint distribution of i.i.d $X_1, \ldots, X_n$ under P. Then note that

$$\rho(P_0^n, P_1^n) = \rho^n(P_0, P_1).$$

Theorem E.30. *$S_{P_0,P_1}(\phi)$ is minimized by taking $\phi = \phi^*$ a.e. μ, where ϕ^* is any test satisfying*

$$\phi^*(x) = \begin{cases} 1 & \text{if } p_1(x) > p_0(x), \\ 0 & \text{if } p_1(x) < p_0(x). \end{cases}$$

Furthermore,

$$S(P_0, P_1) = S_{P_0,P_1}(\phi^*) = 1 - \frac{1}{2}||P_1 - P_0||.$$

Lemma E.31. *Suppose $||P_n - Q_n||_1 \to 0$. Then P_n and Q_n are mutually contiguous. Furthermore, for any sequence of test function ϕ_n,*

$$\int \phi_n \, dP_n - \int \phi_n \, dQ_n \to 0.$$

Theorem E.32. *Suppose*

$$c_1 = \liminf nH^2(P_{\theta_0}, P_{\theta_n}) \leq \limsup nH^2(P_{\theta_0}, P_{\theta_n}) = c_2.$$

Then,

$$1 - [1 - \exp(-2c_2)]^{1/2} \leq \liminf S(P_{\theta_0}^n, P_{\theta_n}^n) \leq \limsup S(P_{\theta_0}^n, P_{\theta_n}^n) \leq \exp(-c_1).$$

Theorem E.33.

1. *If $nH^2(P_n, Q_n) \to 0$, then $||Q_n^n P_n^n||_1 \to 0$ and $\{Q_n^n\}$ are contiguous.*
2. *If $nH^2(P_n, Q_n) \to \infty$, then $S(P_n^n, Q_n^n) \to 0$ and $\{P_n^n\}$ and $\{Q_n^n\}$ are not contiguous.*

E.8 A Convolution Representation Theorem

Definition E.34. With $\theta_n = \theta + \frac{h}{\sqrt{n}}$, the estimate T_n is said to be regular if

$$\mathcal{L}\left[\, \sqrt{n}(T_n - \theta_n) \,|\, P_{n,\theta_n} \,\right] \Rightarrow \theta, \tag{E.28}$$

a probability measure.

Theorem E.35. *Under assumptions* (E.4)*, for regular estimates,*

$$\mathcal{L}(\theta) = \mathcal{N}(0, \Gamma^{-1}(\theta)) * \mathcal{L}^*(\theta), \tag{E.29}$$

where $\mathcal{L}^*(\theta)$ is a specific probability measure.

E.9 LAM [Fabian and Hannan(1982) page 463]

Let γ_n be the central sequence and M_n be a $k \times k$ positive definite matrix such that

$$\|M_n^{-1}\| \to 0, \tag{E.30}$$

and set

$$\theta_{nh}\,(\text{to be shortened to } \theta_n) = \theta + M_n^{-1}h, \quad h \in \mathbb{R}^d \tag{E.31}$$

so that $\theta \in \Theta$ for all sufficiently large n. If Condition LAN $\langle \theta, M_n, \gamma_n \rangle$ holds then $\langle Z_n \rangle$ is LAM(θ) (locally asymptotically minimax at θ) if $\langle Z_n \rangle$ is a sequence of estimates for which

$$\lim_{k\to\infty} \lim_{n\to\infty} \sup_{||M_n^{1/2}(\delta-\theta)||\le K} E_{n,\delta} l(Q_n M_n^{1/2}(Z_n - \delta)) \ge \mathcal{N} l$$

holds for every sequence $\langle Q_n \rangle$ in the collection of all orthogonal $k \times k$ matrices and for every bounded loss function l on $\mathbb{R}^k$.

Definition E.36. A sequence $\langle Z_n \rangle$ of estimates is called *regular*(θ) if

$$M_n^{1/2}(Z_n - \theta) - \gamma_n \to 0 \quad \text{in } \langle E_{n,\theta} \rangle\text{-prob.}$$

Theorem E.37. *Let $\langle Z_n \rangle$ be a sequence of estimates. Then the regularity(θ) of $\langle Z_n \rangle$ implies*

$$E_{n,\delta_n}^{M_n^{1/2}(Z_n - \delta_n)} \Rightarrow \mathcal{N}$$

for every sequence $\delta_n = \theta + M_n^{-1/2} t_n$ such that $\langle t_n \rangle$ is bounded; the latter property, in turn, implies that $\langle Z_n \rangle$ is LAM (θ).

E.10 Discrete sequences of estimators

The discrete sequences of estimators $\{\bar{\theta}_n\}$ satisfies that $\bar{\theta}_n$ is given by one of the vertices of $\{\theta : \theta = n^{-1/2}(i_1, \ldots, i_{p+q}), i_j \in Z\}$ nearest to θ_n, which is a sequence with

$$\sqrt{n}(\theta_n - \theta_0) \quad \text{is bounded by a constant } c > 0.$$

This idea is due to Le Cam(1960), (1969), (1970) for construction of an efficient estimator.

Theorem E.38. *If $\boldsymbol{P} = \{P_\theta; \theta \in \Theta\}$ is a regular parametric model on a Euclidean space $\boldsymbol{X}$ and θ is identifiable, then there exist uniformly $\sqrt{n}$-consistent estimates of θ.*

The steps are as follows:

Step 1 Construct $\tilde{\theta}_n$ uniformly $\sqrt{n}$-consistent as in theorem 2.3 below.

Step 2 Form a grid of cubes with sides of length $cn^{-1/2}$ over $\mathbb{R}^k$, given $\tilde{\theta}_n$, define θ_n^* to be the midpoint of the cube into which $\tilde{\theta}_n$ fallen. (This means that θ_n^* is also uniformly $\sqrt{n}$ consistent.)

Step 3 Define

$$\hat{\theta}_n = \theta_n^* + n^{-1} \sum_{i=1}^{n} I^{-1}(\theta) \dot{\boldsymbol{l}}(X_i, \theta_n^*).$$

Theorem E.39. *If $\boldsymbol{P}$ is a regular parametric model and if there exists a uniformly $\sqrt{n}$-consistent estimator $\tilde{\theta}_n$ of θ, then the estimator $\hat{\theta}_n$ given above is a uniformly efficient estimator of θ.*

note. It is important for the result above that the sample space is Euclidean.
note2. The result is also important since even if the maximum likelihood estimate $\hat{\theta}_n$ does not exist, we can define a one-step Newton-Raphson approximate 'solution' by

$$\hat{\theta}_n^{\text{approx}} = \tilde{\theta}_n + \left[-\frac{1}{n} \sum_{i=1}^{n} \ddot{\boldsymbol{l}}(X_i, \tilde{\theta}_n) \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} \dot{\boldsymbol{l}}(X_i, \tilde{\theta}_n).$$

E.11 LAN for ARMA process in Kreiss (1987)

The LAN property is established for ARMA model by using the assumptions of Roussas(1979). Similar conditions sufficient for the LAN property are given in Swensen(1985).

Theorem E.40 ((K-Theorem 3.1)LAN property for ARMA models)**.** *Let $\{h_n\} \subset \mathbb{R}^{p+q}$ be a bounded sequence and $\theta_n = \theta_0 + n^{-1/2} h_n$. Under our assumptions (A1)–(A4) and (A6) in Chapter 3, we have for*

$$\Delta_n(\theta) = \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(e_j(\theta)) Z(j-1; \theta, \theta), \quad \dot{\varphi} = -f'/2f,$$

the following two results:

$$\log[\, dP_{n,\theta_n} / dP_{n,\theta_0} \,] - h_n^T \Delta_n(\theta_0) + \frac{1}{2} h_n^T I(f) \Gamma(\theta_0) h_n \to 0,$$

in P_{n,θ_0}-probability, where $\Gamma(\theta_0)$ is defined in Theorem 3.5 below (approximation of the log-likelihood ratio).

$$\mathcal{L}(\Delta_n(\theta_0) | P_{n,\theta_0}) \Rightarrow \mathcal{N}(0, I(f)\Gamma(\theta_0)),$$

where "$\Rightarrow$" denotes weak convergence (asymptotic normality of the approximating statistic).

Corollary E.41. Under the same assumption as above $\{P_{n,\theta_0}\}$ and $\{P_{n,\theta_n}\}$ are contiguous in the sense of Definition 2.1, Roussas (1972), page 7, and

$$\mathcal{L}(\Delta_n(\theta_0) - I(f)\Gamma(\theta_0)h_n | P_{n,\theta_n}) \Rightarrow \mathcal{N}(0, I(f)\Gamma(\theta_0)).$$

E.12 The sufficient conditions for local asymptotic normality

The 4 theorems below guarantee that the sufficient conditions for the LAN in Roussas (1979) are fulfilled.

Theorem E.42. *For each $\theta_0 \in \Theta$, the random functions $\phi_j(\theta_0, \cdot)$ are differentiable in q.m. $[\,P_{\theta_0}\,]$ uniformly in $j \geq 1$. That is, there are $(p+q)$-dimensional r.v.'s $\dot{\phi}_j(\theta_0) = \dot{\varphi}(e_j^0)Z(j-1;\theta_0,\theta_0) = \dot{\varphi}(e_j^0)Z^0(j-1)$ [the q.m. derivative of $\phi_j(\theta_0,\theta)$ with respect to θ at θ_0] such that*

$$\frac{\phi_j(\theta_0, \theta_0 + \lambda h) - 1}{\lambda} - h^T \dot{\phi}_j(\theta_0) \to 0, \quad \textit{in q.m. } [P_{\theta_0}] \textit{ as } \lambda \to 0$$

uniformly on bounded sets of $h \in \mathbb{R}^{p+q}$ and uniformly in $j \in \mathbb{N}$. Finally, $\dot{\phi}_j(\theta_0)$ is measurable with respect to $\mathcal{A}_j$.

Theorem E.43. *For each $\theta_0 \in \Theta$ and each $h \in \mathbb{R}^{p+q}$, the sequence $\{(h^T\dot{\phi}_j(\theta_0))^2\}, j \in \mathbb{N}$, is uniformly integrable with respect to P_{θ_0}.*

Theorem E.44. *For each $\theta_0 \in \Theta$ and $j \geq 1$ let the $(p+q) \times (p+q)$-dimensional covariance matrix $\Gamma_j(\theta_0)$ be defined by*

$$\Gamma_j(\theta_0) = 4E_{\theta_0}[\,\dot{\phi}_j(\theta_0)\dot{\phi}_j^T(\theta_0)\,] = I(f)E_{\theta_0}[\,Z(j-1;\theta_0,\theta_0)Z^T(j-1;\theta_0,\theta_0)\,].$$

Then $\Gamma_j(\theta_0) \to \Gamma(\theta_0)I(f)$, as $j \to \infty$, in any one of the standard norms in $\mathbb{R}^{p+q}$, and $\Gamma(\theta_0)$ is positive definite.

Theorem E.45. *(i) For each $\theta_0 \in \Theta$, each $h \in \mathbb{R}^{p+q}$ and for the probability measure P_{θ_0}, the WLLN holds for the sequence $\{[\,h^T\dot{\varphi}_j(\theta_0)\,]^2, j \in \mathbb{N}\}$. Also*
(ii)

$$\frac{1}{n}\sum_{j=1}^{n}\{E_{\theta_0}[\,(h^T\dot{\phi}_j(\theta_0))^2|\mathcal{A}_{j-1}\,] - [\,h^T\dot{\phi}_j(\theta_0)\,]^2\} \to 0, \quad \textit{as } n \to \infty,$$

in P_{θ_0}-probability.

Existence and construction of LAM estimates

Lemma E.46 (K-Lemma 4.1)**.** *Under assumptions (A1)–(A3) and (A6), we have for any sequence $\{Z_n\}$ of estimates the following implication:*

$$\sqrt{n}(Z_n - \theta_0) - \frac{\Gamma(\theta_0)^{-1}}{I(f)}\Delta_n(\theta_0) = o_{P_{\theta_0}}(1) \quad (\{Z_n\} \text{ is called } \theta_0\text{-regular})$$

implies that $\{Z_n\}$ is LAM.

Theorem E.47 (Existence of LAM estimators)**.** *Assume $\{\bar{\theta}_n\} \subset \Theta$ is discrete and $\sqrt{n}$-consistent for $\theta_0 \in \Theta$. Then $\hat{\theta}_n$ defined below is regular:*

$$\hat{\theta}_n = \bar{\theta}_n + \frac{1}{\sqrt{n}}\frac{\hat{\Gamma}_n(\bar{\theta}_n)^{-1}}{I(f)}\Delta_n(\bar{\theta}_n),$$

$$\hat{\Gamma}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} Z(j-1;\theta,\theta)Z^T(j-1;\theta,\theta).$$

Construction of adaptive estimates

Theorem E.48. *Let $\{\bar{\theta}_n\} \subset \Theta$ be a discrete and $\sqrt{n}$-consistent sequence of estimators of θ_0. Under our assumptions (A1)–(A3),(A6)–(A9) and*

$$\tilde{\Delta}_n(\bar{\theta}_n) - \Delta_n(\bar{\theta}_n) = o_{P_{\theta_0}}(1)$$

holds, if $c_n \to \infty$, $g_n \to \infty$, $\sigma(n) \to 0$, $d_n \to 0$, $\sigma(n)c_n \to 0$, $g_n\sigma(n)^{-4}/n \to 0$ and $n\sigma(n)$ stays bounded.

E.13 LAQ

Denote the log-likelihood ratio statistic by

$$\Lambda(\theta + \delta_n \boldsymbol{t}_n; \theta) = \log \frac{dP_{\theta+\delta_n \boldsymbol{t}_n}}{dP_\theta}$$

Definition E.49. The family $\mathcal{P}_n = \{P_{\gamma,n}; \gamma \in \Theta\}$ is termed locally asymptotically quadratic (LAQ) at Θ if there exists a vector $\boldsymbol{S}_n$ and an a.s. non-negative definite matrix $\boldsymbol{K}_n$, both possibly random, such that for every $\boldsymbol{t}_n \in B$,

$$\Lambda(\theta + \delta_n \boldsymbol{t}_n; \theta) - \boldsymbol{t}_n' \boldsymbol{S}_n + \frac{1}{2}\boldsymbol{t}_n' \boldsymbol{K}_n \boldsymbol{t}_n \xrightarrow{p} 0.$$

In particular, if $\boldsymbol{K}_n$ can be taken as non-random, then the LAQ family reduces to LAN family.

Theorem E.50 (Hájek- LeCam-Inagaki theorem)**.** *Consider the LAMN family $\mathcal{F}_{\theta,n} = \{P_{\theta+\delta_n\tau,n}; \tau$ being a p-dimensional vector$\}$, and assume that under $P_{\theta,n}$, the matrx $\boldsymbol{K}_n$ converges in probability to some non-random $\boldsymbol{K}$. For a given non-random matrix $\boldsymbol{A}$, consider the parameter $\boldsymbol{A}\tau$, and let $\boldsymbol{T}_n$ be an estimator of $\boldsymbol{A}\tau$, such that*

$$\boldsymbol{T}_n - \boldsymbol{A}\tau H(\cdot),$$

in distribution under $P_{\theta+\delta_n\tau,n}$, where $H(\cdot)$ does not depend on τ. Then H is the distribution of the random vector $\boldsymbol{A}\boldsymbol{K}^{-1/2}\boldsymbol{Z}+\boldsymbol{U}$, where $\boldsymbol{Z}$ and $\boldsymbol{U}$ are stochastically independent, and $\boldsymbol{Z}$ is asymptotically normal $(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix.

E.14 LAMN

Let δ_n be a $k \times k$ positive definite matrix such that

$$\|\delta_n^{-1}\| \to 0, \tag{E.32}$$

and set

$$\theta_{nh} \,(\text{to be shortened to } \theta_n) = \theta + \delta_n^{-1}h, \quad h \in \mathbb{R}^d \tag{E.33}$$

so that $\theta \in \Theta$ for all sufficiently large n.

Remark E.51. δ_n is a generalized matrix instead of $\sqrt{n}$.

Definition E.52. The sequence of experiments $\{\,(\mathcal{X}, \mathcal{A}_n, P_{n,\theta}); \theta \in \Theta\,\}$, $n \geq 1$, is said to be *Locally Asymptotically Mixed Normal*, if the following two conditions are satisfied:

1. There exists a sequence $\{W_n(\theta)\}$, $n \geq 1$, of $\mathcal{A}_n$-measureble k-dimensional random vectors, and a sequence $\{T_n(\theta)\}$, $n \geq 1$, of $\mathcal{A}_n$-measurable $k \times k$ symmetric and a.s. $[\,P_{n,\theta}\,]$ positive definite random matrices, such that, for every $h \in \mathbb{R}^k$,

$$\Lambda_n(\theta_n, \theta) - \left[h' T_n^{1/2}(\theta) W_n(\theta) - \frac{1}{2} h' T_n(\theta) h \right] \to 0 \quad \text{in } P_{n,\theta}\text{-probability.} \tag{E.34}$$

2. There exists an a.s. $[\,P_{n,\theta}\,]$ positive definite $k \times k$ symmetric random matrix $T(\theta)$, such that

$$\mathcal{L}\{\,[\,W_n(\theta), T_n(\theta)\,] \mid P_{n,\theta}\,\} \Rightarrow \mathcal{L}\{\,[\,W(\theta), T(\theta)\,] \mid P_\theta\,\}, \tag{E.35}$$

   where $W(\theta) \sim \mathcal{N}(0, I_k)$ and is independent of $T(\theta)$.

Remark E.53. By setting,

$$\Delta_n(\theta) = T_n^{1/2}(\theta)W_n(\theta) \quad \text{and} \quad \Delta(\theta) = T^{1/2}(\theta)W(\theta) \tag{E.36}$$

Remark E.54. The points at which LAN of LAMN do not hold are called 'critical points'.

Remark E.55. For LAQ experiments, it is observed that, under the contiguity of $\{P_{n,\theta}\}$ and $\{P_{n,\theta_n}\}$, the following relation holds:

$$\mathcal{E}\left[\exp(h'\Delta - \frac{1}{2}h'Th)\right] = 1 \quad \text{for all } h, \tag{E.37}$$

where Δ and T are as above.

Appendix F: RANK

Consider that we have N samples $(X_1, \ldots, X_N)$.

F.1 Definitions

Ordered Statistics

$$X_{N(1)} \leq \cdots \leq X_{N(N)},$$

which makes a seq in the increasing order is an asymptotic case. If the asymptotic property is not considered, then we may write as follows in the same meaning:

$$X^{(1)} \leq \cdots \leq X^{(N)}.$$

Rank

R will stand for the vector of ranks $(R_1, \ldots, R_N)$. r and $(r_1, \ldots, r_N)$ will be the realization of R, respectively. In the asymptotic case we use

$$R_{Ni}.$$

It is the position number of the i-th sample in N samples. The property is

$$X_i = X_{N(R_{Ni})}.$$

note1. If X_i is tied with some other observations, then we can not define the rank uniquely. In this case, we have two ways to solve this problem as follows:

1. Let $X_i = X_{N(j)}$ for all X_i, where they have the same value, and j is the average of all ranks that these samples take;
2. Let $R_{Ni} = \sum_{j=1}^{N} 1_{\{X_j \le X_i\}}$. (uprank)

However, we will assume the distribution from which the sample is continuous. It means that the case above will be a null set.

note2. the pair $(X^{(i)}, R)$ is a sufficient statistic for any system of distributions determined by densities.

Linear Rank Statistic

$$\sum_{i=1}^{N} a_N(i, R_{Ni})$$

for a given $(N \times N)$ matrix $(a_N(i,j))$. This is the sum of the elements of the matrix $(a_N(i,j))$.

Example 4. Let $X = (2, 3, 1)$. Then

$$\begin{aligned} (1, R_{N1}) &= (1, 2) \\ (2, R_{N2}) &= (2, 3) \\ (3, R_{N3}) &= (3, 1). \end{aligned}$$

Thus, the position of the matrix can be shown as

$$\begin{pmatrix} a_{11} & a_{12} & \bigcirc \\ \bigcirc & a_{22} & a_{23} \\ a_{31} & \bigcirc & a_{33} \end{pmatrix}.$$

Simple Linear Rank Statistics

$$\sum_{i=1}^{N} c_{Ni} a_{N,R_{Ni}};$$

This is a form of the sum of the elements of the matrix multiplied by some coefficients. Comparing the definition of linear rank statistics, the function of i and R_{Ni} is decomposed into two parts–the function of i and the function of R_{Ni}. It also can be seen as the form of a linear combination of the function of ranks.

Coefficients

$$(c_{N1}, \ldots, c_{NN});$$

Scores

$$(a_{N1}, \ldots, a_{NN}).$$

i-th Smallest Coordinate

$$o_i(x),$$

and obviously $x^{(i)} = o_i(x)$.

The System where the Distribution of $(X_1, \ldots, X_N)$ is Symmetric and Determined by a Density

$$p(x_{r_1}, \ldots, x_{r_N}) = p(x_1, \ldots, x_N), \quad r \in R,$$

if and only if $p \in H_*$.

The System where the Observation is iid

$$p = \prod_{i=1}^{N} f(x_i),$$

where f(x) may be an arbitrary one-dimensional density if and only if $p \in H_0$. *note.* $H_0 \subset H_*$.

Incomplete Beta Function Ratio I_z(a,b)

$$F(x) = I_x(a, b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}\, dt}{B(a, b)}, \quad 0 \leq x \leq 1; a, b > 0.$$

The mean of Beta distribution is $\frac{a}{a+b}$, the mode is $\frac{a-1}{a+b-2}$, the variance is $\frac{ab}{(a+b)^2(a+b+1)}$, the coefficient of variation is $\sqrt{\frac{b}{a(a+b+1)}}$, and the skewness is $\frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$.

F.2 Some Lemmas

Lemma F.1. *If X is governed by the density q, then $X^{(\cdot)}$ is governed by the density*

$$\bar{q}(x^{(1)}, \ldots, x^{(N)}) \triangleq \sum_{r \in R} q(x^{(r_1)}, \ldots, x^{(r_N)}), \quad x^{(\cdot)} \in \boldsymbol{X}^{(\cdot)}.$$

Moreover,

$$Q(R = r | X^{(\cdot)} = x^{(\cdot)}) = \frac{q(x^{(r_1)}, \ldots, x^{(r_N)})}{\bar{q}(x^{(1)}, \ldots, x^{(N)})}, \quad r \in R, x^{(\cdot)} \in \boldsymbol{X}^{(\cdot)},$$

holds with Q being the probability distribution corresponding to q.

Proof. For any $A \in \mathcal{A}^{(\cdot)}$, it holds that

$$\begin{aligned}\int \cdots \int_{X^{(\cdot)} \in A} q(x_1, \ldots, x_N) dx_1 \ldots dx_N &= \sum_{r \in R} \int \cdots \int_{X^{(\cdot)} \in A, R=r} q(x_1, \ldots, x_N) dx_1 \ldots dx_N \\ &= \sum_{r \in R} \int \cdots \int_A q(x^{(r_1)}, \ldots, x^{(r_N)}) dx^{(1)} \ldots dx^{(N)}\end{aligned}$$

Note that the Jacobian is 1 in this case. □

note. $\bar{q}$ do not have to be equal to each other. See the following example.

Example 5. Let $\Omega = (1, 2, 3)$ and the probability on it is defined as

$$(p_1, p_2, p_3) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2}).$$

After taking one sample from Ω, we will have the sample set like one of the following three cases:

$$\begin{aligned}\Omega' = (1, 3), &\qquad (p_1, p_3) = (\frac{1}{3}, \frac{2}{3}); \\ \Omega' = (2, 3), &\qquad (p_2, p_3) = (\frac{1}{3}, \frac{2}{3}); \\ \Omega' = (1, 2), &\qquad (p_1, p_2) = (\frac{1}{2}, \frac{1}{2}).\end{aligned}$$

Then the probability for $(x^{(1)}, x^{(2)})$ will be $\frac{5}{12}$, and the probability for $(x^{(2)}, x^{(1)})$ will be $\frac{7}{12}$. Furthermore,

$$\begin{aligned}\bar{q}(1, 2) &= \frac{1}{6}; \\ \bar{q}(2, 3) &= \frac{5}{12}; \\ \bar{q}(1, 3) &= \frac{5}{12}.\end{aligned}$$

This example is a special case, and what we will think next is the property on the system 7.8 and 7.9.

Lemma F.2. *Let $X_1, \ldots, X_N$ be a random sample from a continuous distribution function F with density f. Then*

(1) the vectors $X_{N()}$ and R_N are independent;

(2) the vector $X_{N()}$ has density $N! \prod_{i=1}^{N} f(x_i)$ on the set $x_1 < \cdots < x_N$;

(3) the variable $X_{N(i)}$ has density $N\binom{N-1}{i-1}F(x)^{i-1}(1-F(x))^{N-i}f(x)$; for F the uniform distribution on $[0,1]$, it has mean $i/(N+1)$ and variance $i(N-i+1)/((N+1)^2(N+2))$;

(4) the vector R_N is uniformly distributed on the set of all $N!$ permutations of $1,2,\dots,N$;

(5) for any statistic T and permutation $r=(r_1,\dots,r_N)$ of $1,2,\dots,N$,

$$E(T(X_1,\dots,X_N)|R_N=r)=ET(X_{N(r_1)},\dots,X_{N(r_n)});$$

(6) for any simple linear rank statistic $T=\sum_{i=1}^N c_{Ni}a_{N,R_{Ni}}$,

$$ET=N\bar{c}_N\bar{a}_N;\quad \mathrm{Var}T=\frac{1}{N-1}\sum_{i=1}^N(c_{Ni}-\bar{c}_N)^2\sum_{i=1}^N(a_{Ni}-\bar{a}_N)^2.$$

Proof.

(1)-(4) It is obvious from Lemma 1.

(5) Just change the rotation of the random variables, then we can see it by the virtue of the independence between $X_{N()}$ and R_N.

(6) From $P(R_{Ni}=j)=\frac{1}{N}$,

$$ET=\frac{1}{N}\sum_{i=1}^N c_{Ni}a_{Ni}=N\bar{c}_N\bar{a}_N.$$

The second statement is from a very tedious calculation.

□

note. The sufficient condition for this lemma is very important. Rank statistics are not always distribution-free but when the observations are independent and *identically distributed.*

Corollary F.3. As the same condition, the variable $X_{N(i)}$ has density

$$F_{N(i)}(x)=I_{F(x)}(i,N-i+1)=\frac{N!}{(i-1)!(N-i)!}\int_0^{F(x)}u^{i-1}(1-u)^{N-i}\,du.$$

F.3 The Necessary Condition for Asymptotically Normality

The scores are generated through a given function $\phi : [0,1] \to \mathbb{R}$ in one of two ways. Either

$$a_{Ni} = E\phi(U_{N(i)}), \tag{F.1}$$

where $U_{N(1)}, \ldots, U_{N(N)}$ are the order statistics of a sample of size N from the uniform distribution on $[0,1]$; or

$$a_{Ni} = \phi(\frac{i}{N+1}). \tag{F.2}$$

For well-behaved functions ϕ, these two definitions are closely related and almost identical, since $EU_{N(i)} = \frac{i}{N+1}$.

note. Scores of the first type correspond to the locally most powerful rank tests; scores of the second type are attractive in view of their simplicity.

Examples of Scores

1. The standard normal case: $\phi_f(x) = x$; $(\mathcal{I} = 1)$;
2. The standard logistic case: $\phi_f(x) = \frac{\pi}{\sqrt{3}}(2F(x) - 1)$; $(\mathcal{I} = \frac{\pi^2}{9})$;
3. The standard logistic case: $\phi_f(x) = \sqrt{2}\text{sign}(x)$; $(\mathcal{I} = 2)$.

Before look into the theorem on the rank, first we introduce a useful theorem.

Theorem F.4. *Let S_n be linear spaces of random variables with finite second moments that contain the constants. Let T_n be random variables with projections $\hat{S}_n$ onto $\mathcal{S}_n$. If $\text{Var}T_n/\text{Var}\hat{S}_n \to 1$ then*

$$\frac{T_n - ET_n}{\text{sd}T_n} - \frac{\hat{S}_n - E\hat{S}_n}{\text{sd}\hat{S}_n} \xrightarrow{p} 0.$$

Proof. Take the expectation of the left side term, then we can see that it is 0.

Next think about the variance about the left side term.

$$\begin{aligned}
\mathrm{Var}\left[\frac{T_n - ET_n}{\mathrm{sd}T_n} - \frac{\hat{S}_n - E\hat{S}_n}{\mathrm{sd}\hat{S}_n}\right] &= 2 - 2\mathrm{Cov}\left(\frac{T_n - ET_n}{\mathrm{sd}T_n}, \frac{\hat{S}_n - E\hat{S}_n}{\mathrm{sd}\hat{S}_n}\right) \\
&= 2 - 2\frac{\mathrm{Cov}(T_n, \hat{S}_n)}{\mathrm{sd}T_n \mathrm{sd}\hat{S}_n} \\
&= 2 - 2\frac{\mathrm{Var}\hat{S}_n}{\mathrm{sd}T_n \mathrm{sd}\hat{S}_n} \\
&\to 0,
\end{aligned}$$

which shows the convergence in second mean. □

According to Theorem E.4., we have the following theorem:

Theorem F.5. *Let R_N be the rank vector of an i.i.d. sample $X_1, \ldots, X_N$ from the continuous distribution function F. Let the scores a_N be generated according to* (F.1) *for a measurable function ϕ that is not constant almost everywhere, and satisfies $\int_0^1 \phi^2(u)\,du < \infty$. Define the variables*

$$T_N = \sum_{i=1}^{N} c_{Ni} a_{N,R_{Ni}}, \quad \tilde{T}_N = N\bar{c}_N \bar{a}_N + \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)\phi(F(X_i)).$$

Then the sequences T_N and $\tilde{T}_N$ are asymptotically equivalent in the sense that

$$ET_N = E\tilde{T}_N$$

and

$$\frac{\mathrm{Var}(T_N - \tilde{T}_N)}{\mathrm{Var}T_N} \to 0.$$

The same is true if the scores are generated according to (F.2) *for a function ϕ that is continuous and almost everywhere, is nonconstant, and satisfies*

$$\frac{1}{N}\sum_{i=1}^{N} \phi^2(\frac{i}{N+1}) \to \int_0^1 \phi^2(u)\,du < \infty.$$

Proof. Set $U_i = F(X_i)$, then from Lemma E.2(5), it is seen that

$$E(\phi(U_i)|R_N) = E(\phi(U_{N,R_{Ni}})) = a_{N,R_{Ni}} \quad \text{a.s.}$$

This implies that

$$E(\tilde{T}_N|R_N) = N\bar{c}_N \bar{a}_N + \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N) a_{N,R_{Ni}} = T_N$$

in almost surely mean. If, from Lemma E.2(6),

$$\frac{\mathrm{Var}T_N}{\mathrm{Var}\tilde{T}_N} = \frac{N}{N-1}\frac{\mathrm{Var}\, a_{N,R_{N1}}}{\mathrm{Var}\,\phi(U_1)} \to 1,$$

then the conclusion is obtained by the theorem E.4.

In fact,

$$E(a_{N,R_{N1}} - \phi(U_1))^2 \to 0,$$

which implies the convergence above. This statement is shown by martingale convergence theorem, since each rank vector R_{j-1} is a function of the next rank vector R_j, it is also true that

$$a_{N,R_{N1}} = E(\phi(U_1)|R_1, \ldots, R_N) \quad \text{a.s.}$$

Square-integrability implies the uniformly integrability, which further implies that the martingale convergence theorem is true. We omit the proof that $\phi(U_1)$ is measurable w.r.t R_1, R_2, ... although it is also important to show

$$E\phi(U_1) = E(\phi(U_1)|R_1, \ldots).$$

The second statement is shown as follows: Set $b_N i = \phi(\frac{i}{N+1})$, and show the statistics generated by a_{Ni} and b_{Ni} are asymptotically equivalent. By the theorem E.4.,

$$\frac{R_{N1}}{N+1} \to U_1 \quad \text{in quadratic mean},$$

since

$$E(U_1 \,|\, R_{N1}) = \frac{R_{N1}}{N+1},$$

and

$$\mathrm{Var}\frac{R_{N1}}{N+1} \to \frac{1}{12} = \mathrm{Var}U_1.$$

This implies

$$\frac{1}{N}\sum_{i=1}^{N}(a_{Ni} - b_{Ni})^2 \to 0,$$

which shows

$$\frac{\mathrm{Var}(S_N - T_N)}{\mathrm{Var}\, T_N} = \frac{\sum_{i=1}^N (a_{Ni} - b_{Ni} - (\bar{a}_N - \bar{b}_N))^2}{\sum_{i=1}^N (a_{Ni} - \bar{a}_N)^2} \to 0,$$

which completes the proof. □

Theorem F.6 (Lindeberg-Feller central limit theorem)**.** *For each n let $Y_{n,1}, \ldots, Y_{n,k_n}$ be independent random variables. Suppose $E(Y_{n,i}) = 0$, $E(Y_{n,i}^2) = \sigma_{n,i}^2 < \infty$ and $s_n^2 = \sum_{i=1}^{k_n} \sigma_{n,i}^2$. If*

$$\frac{1}{s_n^2} \sum_{i=1}^{k_n} E[\,Y_{n,i}^2 1\{|Y_{n,i}| > \epsilon s_n\} \to 0, \quad \textit{for every } \epsilon > 0,$$

Then

$$\sum_{i=1}^{k_n} \frac{Y_{n,i}}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

Theorem F.7 (a special central limit theorem)**.** *Let Y_1, Y_2, $\ldots$ be independent copies of a random variable with finite expectation μ and finite variance σ^2. Put*

$$T_a = \sum_{i=1}^{N} a_i Y_i, \quad a \in A.$$

Then for

$$\frac{\max_{1 \le i \le N} a_i^2}{\sum_{i=1}^{N} a_i^2} \to 0,$$

the statistics T_a are asymptotically normal (μ_a, σ_a^2) with

$$\mu_a = \mu \sum_{i=1}^{N} a_i, \quad \sigma_a^2 = \sigma^2 \sum_{i=1}^{N} a_i^2.$$

The theorem above is an extension of Lindeberg-Feller central limit theorem. Note that the sufficient condition in the rank statistics case is

$$\frac{\max_{1 \le i \le N} (c_{Ni} - \bar{c}_N)^2}{\sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2} \to 0. \tag{F.3}$$

Proof. From theorem E.6., we only have to show the Lindeberg condition

$$\sigma_a^{-2} \sum_{i=1}^{N} \int_{|x| > \epsilon \sigma_a} x^2 \, dP(a_i(Y_i - \mu) \le x) \to 0.$$

To see this,

$$\begin{aligned} \int_{|x| > \epsilon \sigma_a} x^2 \, dP(a_i(Y_i - \mu) \le x) &= a_i^2 \int_{|y a_i| > \epsilon \sigma_a} y^2 dP(Y_i - \mu \le y) \\ &\le a_i^2 \int_{|y| > \epsilon \sigma \nu_a} y^2 dP(Y_i - \mu \le y), \end{aligned}$$

where

$$\nu_a^2 = \sum_{i=1}^{N} a_i^2 / \max_{1 \le i \le N} a_i^2.$$

Hence

$$\sigma_a^{-2} \sum_{i=1}^{N} \int_{|x| > \epsilon \sigma_a} x^2 \, dP(a_i(Y_i - \mu) \le x) \le \sigma^{-2} \int_{|y| > \epsilon \sigma \nu_a} y^2 \, dP(Y_1 - \mu \le y) \to 0.$$

□

Corollary F.8. If the vector of coefficients c_N satisfies (F.3), and the scores are generated according to (F.1) for a measurable, nonconstant, square-integrable function ϕ, then the sequence of standardized rank statistics

$$\frac{(T_N - ET_N)}{\mathrm{sd} T_N} \xrightarrow{d} \mathcal{N}(0, 1).$$

The same is true if the scores are generated by (F.2) for a function ϕ that is continuous almost everywhere, is nonconstant, and satisfies

$$\frac{1}{N} \sum_{i=1}^{N} \phi^2 \left(\frac{i}{N+1}\right) \to \int_0^1 \phi^2(u) \, du.$$

The next central limit theorem is a case for dependent random variables.

Theorem F.9 (Dependent central limit theorem)**.**
Suppose $T_n = \prod_{j=1}^{k_n} (1 + itX_{n,j})$ *and* $S_n = \sum_{i=1}^{k_n} X_{n,i}$.
Assume for all real t*,*

1. $ET_n \to 1$,
2. $\{T_n\}$ *is uniformly integrable,*
3. $\sum_j X_{n,j}^2 \xrightarrow{p} 1$,
4. $\max_{j \le k_n} |X_{n,j}| \xrightarrow{p} 0$.

Then,

$$S_n \xrightarrow{d} \mathcal{N}(0, 1).$$

note. Note that

$$e^{ix} = (1 + ix) \exp\{-\frac{x^2}{2} + r(x)\}.$$

A generalized CLT for rank is given as follows. The proof is quite involved and we do not show it in this Thesis.

Theorem F.10 (Rank central limit theorem). *Let* $T_N = \sum c_{Ni} a_{N,R_{Ni}}$ *be the simple linear rank statistic with coefficients and scores such that*

$$\max_{1 \le i \le N} \frac{|a_{Ni} - \bar{a}_N|}{\sum_{i=1}^{n} (a_{Ni} - \bar{a}_N)^2} \to 0;$$
$$\max_{1 \le i \le N} \frac{|c_{Ni} - \bar{c}_N|}{\sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2} \to 0.$$

Let the rank vector R_N *be uniformly distributed on the set of all* $N!$ *permutations of* $\{ 1, 2, \ldots, N \}$. *Then the sequence*

$$\frac{T_N - ET_N}{\mathrm{sd} T_N} \xrightarrow{d} \mathcal{N}(0, 1),$$

if and only if, for every $\epsilon > 0$,

$$\sum\sum_{(i,j): \sqrt{N} |a_{Ni} - \bar{a}_N| |c_{Ni} - \bar{c}_N| > \epsilon A_N C_N} \frac{|a_{Ni} - \bar{a}_N|^2 |c_{Ni} - \bar{c}_N|^2}{A_N^2 C_N^2} \to 0,$$

where

$$A_N^2 = \sum_{i=1}^{n} (a_{Ni} - \bar{a}_N)^2, \quad C_N^2 = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2.$$

Bibliography

P.J. Bickel, C.A.J. Klaassen, J.A.W. Ya'acov Ritov, and J.A. Wellner. *Efficient and adaptive estimation for semiparametric models*, volume 1. Springer New York, 1998.

P. Billingsley. *Convergence of probability measures*, volume 493. Wiley-Interscience, 1968.

D.R. Brillinger. *Time Series: Data Analysis and Theory.* Applied Mathmatics, 2001.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, 1991.

S.U. Can, T. Mikosch, and G. Samorodnitsky. Weak convergence of the function-indexed integrated periodogram for infinite variance processes. *Bernoulli*, 16(4): 995–1015, 2010.

R. Davis and S. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, pages 179–195, 1985.

R. Davis and S. Resnick. Limit theory for the sample covariance and correlation functions of moving averages. *The Annals of Statistics*, pages 533–558, 1986.

R.A. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23(2):879–917, 1995.

R.A. Davis, J.E. Marengo, and S.I. Resnick. Extremal properties of a class of multivariate moving averages. 1986.

V. Fabian and J. Hannan. On estimation and adaptive estimation for locally asymptotically normal families. *Probability Theory and Related Fields*, 59(4): 459–478, 1982.

B. Garel and M. Hallin. Local asymptotic normality of multivariate arma processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, 47(3): 551–579, 1995.

Philip S Griffin and David M Mason. On the asymptotic normality of self-normalized sums. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 109, pages 597–610. Cambridge Univ Press, 1991.

J. Hájek, Z. Šidák, and P.K. Sen. *Theory of rank tests.* Academic press New York, 1967.

P. Hall and C.C. Heyde. *Martingale limit theory and its application.* Academic press New York, 1980.

M. Hallin and M.L. Puri. Rank tests for time series analysis a survey. 1991.

M. Hallin and B.J.M. Werker. Optimal testing for semi-parametric ar models-from gaussian lagrange multipliers to autoregression rank scores and adaptive tests. *statistics textbooks and monographs*, 158:295–350, 1999.

M. Hallin and B.J.M. Werker. Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9(1):137–165, 2003.

M. Hallin, J.F. Ingenbleek, and M.L. Puri. Linear serial rank tests for randomness against arma alternatives. *The Annals of Statistics*, pages 1156–1181, 1985.

M. Hallin, J.F. Ingenbleek, and M.L. Puri. Linear and quadratic serial rank tests for randomness against serial dependence. *Journal of time series analysis*, 8(4): 409–424, 1987.

P.J. Huber. *Robust statistics.* Wiley, New York, 2009.

C. Klüppelberg and T. Mikosch. Spectral estimates and stable processes. *Stochastic processes and their applications*, 47(2):323–344, 1993.

C. Klüppelberg and T. Mikosch. The integrated periodogram for stable processes. *The Annals of Statistics*, pages 1855–1879, 1996.

J.P. Kreiss. On adaptive estimation in stationary arma processes. *The Annals of Statistics*, pages 112–133, 1987.

E.L. Lehmann and J.P. Romano. *Testing statistical hypotheses.* Springer, 2005.

B.F. Logan, CL Mallows, SO Rice, and LA Shepp. Limit distributions of self-normalized sums. *The Annals of Probability*, 1(5):788–809, 1973.

H. Lütkepohl. *New introduction to multiple time series analysis*. Cambridge Univ Press, 2005.

Wilhelm Magnus and Fritz Oberhettinger. *Formulas and theorems for the functions of mathematical physics*. Chelsea Pub. Co., 1954.

M.M. Meerschaert and H.P. Scheffler. *Limit distributions for sums of independent random vectors: Heavy tails in theory and practice*, volume 321. Wiley-Interscience, 2001.

Resnick. S. Mikosch, T. and G. Samorodnitsky. The maximum of the periodogram for a heavy-tailed sequence. *The Annals of Probability*, 28(2):885–908, 2000.

T. Mikosch, T. Gadrich, C. Kluppelberg, and R.J. Adler. Parameter estimation for arma models with infinite variance innovations. *The Annals of Statistics*, 23 (1):305–326, 1995.

J.P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, 2012.

H. Ogata and M. Taniguchi. An empirical likelihood approach for non-gaussian vector stationary processes and its application to minimum constrast estimation. *Australian and New Zealand Jounal of Statistics*, 52(4):451–468, 2010.

A. Owen. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120, 1990.

A.B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

Victor H Peña, Tze Leung Lai, and Qi-Man Shao. *Self-normalized processes: Limit theory and Statistical Applications*. Springer, 2009.

S.I. Resnick. *Extreme values, regular variation, and point processes*. Springer, 2007.

G. Samoradnitsky and MS Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall/CRC, 1994.

A.R. Swensen. The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis*, 16(1): 54–70, 1985.

M. Taniguchi and J. Hirukawa. Generalized information criterion. *Journal of Time Series Analysis*, 33(2):287–297, 2011.

M. Taniguchi and Y. Kakizawa. *Asymptotic theory of statistical inference for time series*. New York: Springer-Verlag, 2000.

A.W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Christos Xenophontos. A formula for the nth derivative of the quotient of two functions. 2007.