2. G_2-ESTIMATOR OF THE REAL STIELTJES TRANSFORM OF THE NORMALIZED SPECTRAL FUNCTION OF COVARIANCE MATRICES

Consider the main problem of the statistical analysis of observations of large dimension: the estimation of Stieltjes' transforms of the normalized spectral functions

$$\mu_{m_n}(x) = m_n^{-1} \sum_{k=1}^{m_n} \chi(\lambda_k < x)$$

of the covariance matrices R_{m_n} from the observations of the random vector $\vec{\xi}$ with the covariance matrix R_{m_n}, where λ_k are eigenvalues of matrix R_{m_n}. Note that many analytic functions of the covariance matrices that are used in multivariate statistical analysis can be expressed through the spectral function $\mu_{m_n}(x)$. For example, the function

$$m_n^{-1}\mathrm{Tr} f(R_{m_n}) = \int_0^\infty f(x)\,\mathrm{d}\mu_{m_n}(x),$$

where $f(x)$ is an analytical function.

The function

$$\varphi(t, R_{m_n}) = \int_0^\infty (1+tx)^{-1}\,\mathrm{d}\mu_{m_n} = m_n^{-1}\mathrm{Tr}(I + tR_{m_n})^{-1},\ t > 0,$$

is called Stieltjes' transform of the function $\mu_{m_n}(x)$. A consistent estimator of Stieltjes' transform $\varphi(t, R_{m_n})$ is equal to: $G_2(t, \hat{R}_{m_n}) = \varphi(\hat{\theta}_n(t), \hat{R}_{m_n})$, where $\hat{\theta}_n(t)$ is the positive solution of the equation

$$\theta(1 - m_n(n-1)^{-1} + m_n(n-1)^{-1}\varphi(\theta, \hat{R}_{m_n})) = t,\ t \geq 0.$$

It is obvious that the positive solution of this equation exists and is unique as $t \geq 0,\ m_n(n-1)^{-1} < 1$.

Let the independent observations $\vec{x}_1, ..., \vec{x}_n$ of the m_n-dimensional random vector $\vec{\xi}$ be given. Assume that the G-condition is fulfilled:

$$\limsup_{n\to\infty} m_n n^{-1} < 1,\quad 0 < c_1 \leq \lambda_i \leq c_2 < \infty,\ i = 1, ..., m_n,$$

and let the components of the vector $(\eta_{1k}, ..., \eta_{m_n k})^{\mathrm{T}} = R_{m_n}^{-1/2}(\vec{\xi} - \mathbf{E}\vec{\xi})$ be independent, and

$$\sup_n \sup_{k=1,...,n} \sup_{i=1,...,m_n} \mathbf{E}|\eta_{ik}|^{4+\delta} < \infty,\quad \delta > 0.$$

Then ([Gir38–41], [Gir43–45], [Gir54], [Gir58], [Gir69], [Gir76], [Gir84])

$$\lim_{n\to\infty} \mathbf{P}\left\{[G_2(t, \hat{R}_{m_n}) - \varphi(t, R_{m_n})]\sqrt{(n-1)m_n}\, a_n(t) + c_n(t) < x\right\}$$
$$= (2\pi)^{-1/2} \int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\mathrm{d}y,$$

as $t > 0$, where $a_n(t)$ and $c_n(t)$ are some bounded functions.

2.1. G_2-estimator of a complex Stieltjes transform of the normalized spectral function of covariance matrices

Here, the $G_2(z)$-consistent estimator for the trace of the resolvent of covariance matrices (Stieltjes' transform)

$$m_n^{-1}\mathrm{Tr}\left(\hat{R}_{m_n}-zI_{m_n}\right)^{-1},\ z=t+\mathrm{i}s,\ s>0$$

is given as

$$G_2(z)=z^{-1}\hat{\theta}(z)\,m_n^{-1}\mathrm{Tr}\left\{\hat{R}_{m_n}-\hat{\theta}(z)\,I_{m_n}\right\}^{-1},$$

where $\hat{\theta}(z)$ is the measurable complex solution of the equation

$$\hat{\theta}(z)\,\frac{1}{n}\mathrm{Tr}\left\{\hat{R}_{m_n}-\hat{\theta}(z)\,I_{m_n}\right\}^{-1}-\left(1-\frac{m_n}{n}\right)+\frac{\hat{\theta}(z)}{z}=0.$$

THEOREM 2.1. [Gir45] *Suppose that* $\vec{x}_1,\dots,\vec{x}_n$ *is a random vector sample,*

$$\vec{x}_k=R_{m_n}^{1/2}\vec{\xi}_k+\vec{a},\ \ \mathbf{E}\,\vec{\xi}_k=0,\ \mathbf{E}\,\vec{\xi}_k\vec{\xi}_k^T=I_{m_n},\ \ \vec{\xi}_k^T=\{\xi_{ik},i=1,...,m_n\},$$

for any positive defined matrix A_m *whose eigenvalues are bounded by a certain constant*

$$\lim_{n\to\infty}\max_{k=1,\dots,n}n^{-1}\mathbf{E}\left|(\vec{x}_k-\vec{a})^T A\,(\vec{x}_k-\vec{a})-n^{-1}\mathrm{Tr}\,R_{m_n}A\right|=0,$$

$$\lambda_i(R_{m_n})<c_1<\infty,\ \ i=1,...,m_n,$$

$$\liminf_{n\to\infty}m_n n^{-1}>0,\quad \limsup_{n\to\infty}m_n n^{-1}<\infty.$$

Then with probability one for every $S>0$ *and* $T>0$

$$\lim_{n\to\infty}\sup_{\substack{0<c\le \mathrm{Im}z\le S,\\ |\mathrm{Re}z|\le T}}\left|G_2(z)-m_n^{-1}\mathrm{Tr}\left\{R_{m_n}-zI_{m_n}\right\}^{-1}\right|=0,$$

for some $c>0$.

2.2. Modified G_2-estimator

Thus, under some conditions,

$$\lim_{n\to\infty}\sup_{\substack{0<c\le \mathrm{Im}z\le S,\\ |\mathrm{Re}z|\le T}}\left|G_2(z)-m_n^{-1}\mathrm{Tr}\left\{R_{m_n}-zI_{m_n}\right\}^{-1}\right|=0,$$

where $c>0$ is a certain constant (which usually is not small). However, we need to know the trace of the resolvent of the covariance matrix for all $s>0$. Since function $m_n^{-1}\mathrm{Tr}\left\{R_{m_n}-zI_{m_n}\right\}^{-1}$ is analytical in z, $\mathrm{Im}z>0$, we can use many methods for its analytical continuation. For example we can use the Fourier transform and consider the following modified G_2 estimator:

$$G_2(A,B,u+\mathrm{i}v)=\mathrm{i}\int_0^B\left\{\frac{\mathrm{e}^{|sp|}}{\pi}\int_{-A}^{A}\mathrm{Im}G_2(z)\,\mathrm{e}^{-\mathrm{i}tp}\mathrm{d}t\right\}\mathrm{e}^{-p(v-\mathrm{i}u)}\mathrm{d}p,\ \ v>0,$$

where $s > c > 0$.

It is easy to prove that the following assertion is valid:

THEOREM 2.2 [Gir45] *If the conditions of Theorem 2.1 are fulfilled, then with probability one, for every $\varepsilon > 0$,*

$$\lim_{B\to\infty}\lim_{A\to\infty}\lim_{n\to\infty}\sup_{v,\,0<\varepsilon\le u}\left|G_2\left(A,B,u+\mathrm{i}v\right)-m_n^{-1}\mathrm{Tr}\left\{R_{m_n}-\left(u+\mathrm{i}v\right)I_{m_n}\right\}^{-1}\right|=0.$$

2.3. G_2-estimator for the trace of the resolvent of empirical covariance matrix when Lindeberg's condition is not fulfilled

Let $\vec{x}_1,\dots,\vec{x}_n$ be the sample of independent observations of a random vector,

$$\vec{x}_k=R_{m_n}^{1/2}\vec{\xi}_k+\vec{a},\ \ \mathbf{E}\,\vec{\xi}_k=0,\ \ \mathbf{E}\,\vec{\xi}_k\vec{\xi}_k^T=I_{m_n},\ \ \vec{\xi}_k^T=\left\{\beta_k\xi_{ik},i=1,...,m_n\right\},$$

where β_k are independent and do not depend on variables ξ_{ik}.

For this case, the G-equation for the trace of resolvent has the following form [Gir69]

$$b(z)=\mathbf{E}\,m_n^{-1}\mathrm{Tr}\left[\hat{R}_{m_n}-zI_{m_n}\right]^{-1}=m_n^{-1}\sum_{p=1}^{m_n}\frac{1}{\lambda_p n^{-1}\sum\limits_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q(z)+1}-z}+\varepsilon_n,$$

where $\lim_{n\to\infty}\varepsilon_n=0$ and $q(z)$ is satisfies the equation

$$q(z)=m_n^{-1}\sum_{p=1}^{m_n}\frac{\lambda_p}{\lambda_p n^{-1}\sum\limits_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q(z)+1}-z},\quad z=t+\mathrm{i}s,\quad \gamma=\frac{m_n}{n}.$$

Let us express function $q(z)$ through function $b(z)$. One has

$$\begin{aligned}q(z)n^{-1}\sum_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q\left(z\right)+1}&=m_n^{-1}\sum_{p=1}^{m_n}\frac{\lambda_p n^{-1}\sum\limits_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q(z)+1}}{\lambda_p n^{-1}\sum\limits_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q(z)+1}-z}\\&=1+m_n^{-1}\sum_{p=1}^{m_n}\frac{z}{\lambda_p n^{-1}\sum\limits_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q(z)+1}-z}\\&=1+zb\left(z\right).\end{aligned}$$

Hence, in this case, the G_2- estimator has the following form

$$G_2(z)=\hat{b}_n(\theta(z)z)\theta(z),$$

where $\theta(z)$ is any measurable solution of the equation

$$n^{-1}\sum_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2q\left(\theta(z)z\right)+1}=\theta(z)$$

and $q(z)$ is any measurable solution of the equation

$$q(z)n^{-1}\sum_{i=1}^{n}\mathbf{E}\,\frac{\beta_i^2}{\gamma\beta_i^2 q\left(z\right)+1}=1+z\hat{b}_n\left(z\right),$$

with

$$\hat{b}_n\left(z\right)=n^{-1}\mathrm{Tr}\left[\hat{R}_{m_n}-Iz\right]^{-1}.$$

Random Matrices. Kiev University Publishing, Ukraine, 1975, 448pp. (in Russian).

Theory of Random Determinants. Kiev University Publishing, Ukraine, 1975, 368pp. (in Russian).

Limit Theorems for Functions of Random Variables. "Higher School" Publishing, Kiev, Ukraine, 1983, 207pp. (in Russian).

Multidimensional Statistical Analysis. "Higher School" Publishing, Kiev, Ukraine, 1983, 320pp. (in Russian).

Spectral Theory of Random Matrices. "Science" Publishing, Moscow, Russia, 1988, 376pp. (in Russian).

Theory of Random Determinants. Kluwer Academic Publishers, The Netherlands, 1990, 677pp.

Theory of Systems of Empirical Equations. "Lybid" Publishing, Kiev, Ukraine, 1990, 264pp. (in Russian).

Statistical Analysis of Observations of Increasing Dimensions. Kluwer Academic Publishers, The Netherlands, 1995, 286pp.

Theory of Linear Algebraic Equations with Random Coefficients. Allerton Press, Inc, New York, U.S.A. 1996, 320pp.

An Introduction to Statistical Analysis of Random Arrays. VSP, Utrecht, The Netherlands. 1998, 673pp.

Theory of Stochastic Canonical equations. Volume I and II. Kluwer Academic Publishers, (xxiv + 497, xviii + 463).(2001).