Probab. Theory Relat. Fields 131, 528–552 (2005)
Digital Object Identifier (DOI) 10.1007/s00440-004-0384-5

Zhidong Bai · Tailen Hsing

The broken sample problem

Dedicated to Professor Xiru Chen on His 70th Birthday

Received: 20 February 2002 / Revised version: 16 June 2004
Published online: 12 September 2004 –

Abstract. Suppose that $(X_i, Y_i), i = 1, 2, \dots, n$, are iid. random vectors with uniform marginals and a certain joint distribution F_ρ, where ρ is a parameter with $\rho = \rho_o$ corresponds to the independence case. However, the X's and Y's are observed separately so that the pairing information is missing. Can ρ be consistently estimated? This is an extension of a problem considered in DeGroot and Goel (1980) which focused on the bivariate normal distribution with ρ being the correlation. In this paper we show that consistent discrimination between two distinct parameter values ρ_1 and ρ_2 is impossible if the density f_ρ of F_ρ is square integrable and the second largest singular value of the linear operator $h \to \int_0^1 f_\rho(x, \cdot)h(x)dx,\ h \in L^2[0, 1]$, is strictly less than 1 for $\rho = \rho_1$ and ρ_2. We also consider this result from the perspective of a bivariate empirical process which contains information equivalent to that of the broken sample.

1. Introduction

Consider a family of bivariate distributions with a parameter ρ and let F_ρ be the joint cdf. One can think of ρ as a measure of association such as the correlation. We assume that the parameter space contains a specific value ρ_o which corresponds to the independence of the marginals. Let $(X_i, Y_i), i = 1, 2, \dots, n$, be iid. random vectors from this distribution. However, we assume an incomplete or "broken" sample in which the X's and Y's are observed separately, and the information on the pairing of the two sets of observations is lost. Our goal is to investigate the consistent discrimination of the F_ρ, where consistency in this paper refers to weak consistency. In DeGroot and Goel (1980), the problem of estimating the correlation of a bivariate normal distribution based on a broken sample was considered. They showed that the Fisher information at $\rho = 0$ is equal to 1 for all sample

Z. Bai: North East Normal University, China and Department of Statistics and Applied Probability, National University of Singapore, Singapore.
e-mail: stabaizd@leonis.nus.edu.sg. Research supported by NSFC Grant 201471000 and the NUS Grant R-155-000-040-112.

T. Hsing: Texas A&M University and Department of Statistics, Texas A&M University, College Station, Texas, USA.
e-mail: thsing@stat.tamu.edu. Research supported by the Texas Advanced Research Program.

Mathematics Subject Classification (2000): primary: 60F99, 62F12

*Key words or phrases:*Consistent estimation – Empirical process – Gaussian process – Kulback-Leibler information

sizes, which leads to the conjecture that consistent estimation is not possible (if the parameter space contains a neighborhood of 0). However, they failed to give a definitive conclusion.

Since the marginal distributions can be consistently estimated with the broken sample, in order for the problem stated here to make sense we need for ρ to be either not present, or at least not identifiable, in the marginal distributions. With that consideration in mind, we assume without loss of generality that the marginal distributions are uniform [0, 1], for we may otherwise consider $(F_X(X_i), F_Y(Y_i))$ where F_X and F_Y are the marginal distributions of X and Y respectively. Thus, the distribution under ρ_o is the uniform distribution on $[0, 1] \times [0, 1]$. The main purpose of this paper is to try to understand whether it is possible to consistently discriminate two distinct parameter values ρ_1 and ρ_2 based on the broken sample, that is, whether there exists a sequence of statistics T_n of the broken sample, where n refers to the sample size, taking values in $\{\rho_1, \rho_2\}$ and such that

$$\lim_{n\to\infty} P_{\rho_i}(T_n = \rho_i) = 1,\ i = 1, 2. \tag{1}$$

Here and in the sequel, P_ρ denotes probability computation when the true parameter is ρ. The condition under which consistent discrimination rules do not exist turns out to be remarkably simple. Let f_ρ be the density of F_ρ. We will show in Theorem 1 that ρ_1 and ρ_2 cannot be consistently discriminated if for $\rho = \rho_1$ and ρ_2, $\int_0^1 \int_0^1 f_\rho^2(x, y)dxdy < \infty$ and the second largest singular value of the linear operator $h \to \int_0^1 f_\rho(x, \cdot)h(x)dx,\ h \in L^2[0, 1]$, is strictly less than 1.

To give some insight into this result, we consider the two-dimensional empirical process

$$F_n(x, y) = n^{-1}\left(\sum_{i=1}^n I_{\{X_i\le x\}}, \sum_{i=1}^n I_{\{Y_i\le y\}}\right),\ x, y \in [0, 1],$$

which contains all the existing information in the broken sample. It is straightforward to verify that the standardized empirical process $\boldsymbol{Z}_n(x, y) = n^{1/2}(F_n - EF_n)$ converges weakly to a Gaussian process $\boldsymbol{Z} = (Z_1, Z_2)$ in the space $D[0, 1] \times D[0, 1]$ where the Z_i are marginally Brownian bridge with

$$\begin{aligned}\operatorname{cov}(Z_1(x_1), Z_1(x_2)) &= x_1 \wedge x_2 - x_1x_2\\ \operatorname{cov}(Z_2(y_1), Z_2(y_2)) &= y_1 \wedge y_2 - y_1y_2\\ \operatorname{cov}(Z_1(x), Z_2(y)) &= F_\rho(x, y) - xy.\end{aligned}$$

Let $\mathcal{P}_\rho$ henceforth denote the probability distribution of the limiting Gaussian process $\boldsymbol{Z}$ described above under parameter value ρ. Note that the standardization does not involve ρ, so it is reasonable to argue that most of the information about ρ in F_n carries over to $\boldsymbol{Z}$. We also remark in passing that the weak convergence implies that ρ is identifiable in the broken sample setting so long as it is identifiable in the bivariate distribution F_ρ. Suppose that for two given parameter values ρ_1 and ρ_2, $\mathcal{P}_{\rho_1}$ and $\mathcal{P}_{\rho_2}$ are equivalent, also called mutually absolutely continuous and denoted by $\mathcal{P}_{\rho_1} \equiv \mathcal{P}_{\rho_2}$ here. Then it is clearly not possible to discriminate between the two models with probability one based on $\boldsymbol{Z}$. Theorem 3 shows that the same conditions of Theorem 1 plus some additional minor regularity condition ensure that $\mathcal{P}_{\rho_i} \equiv \mathcal{P}_{\rho_o}, i = 1, 2$ and hence $\mathcal{P}_{\rho_1} \equiv \mathcal{P}_{\rho_2}$.

To demonstrate the results, we will revisit the bivariate normal problem in DeGroot and Goel (1980) and show that consistent discrimination of any two bivariate normal distributions with different correlations in $(-1, 1)$ is impossible. We will also consider other examples for which ρ can be consistently discriminated or even estimated.

2. Main results and examples

We assume that F_ρ has a density f_ρ, and write

$$A(\rho) = \int_0^1 \int_0^1 f_\rho^2(x, y)dxdy.$$

Define the linear operator

$$T_\rho : h \to \int_0^1 f_\rho(x, \cdot)h(x)dx, \; h \in L^2[0, 1].$$

Suppose $A(\rho) < \infty$. Then T_ρ is a Hilbert-Schmidt operator and admits the singular-value decomposition (cf. Riesz and Sz.-Nagy, 1955). Since 1 is necessarily a singular value of T_ρ with singular value functions equal to the constant function 1, the singular-value decomposition can be written as

$$T_\rho = 1 + \sum_{i=1}^{\infty} \lambda_{i,\rho} \psi_{i,\rho} \otimes \phi_{i,\rho},$$

where, with

$$(\psi_{i,\rho} \otimes \phi_{i,\rho})h = \left(\int_0^1 \psi_{i,\rho}(x)h(x)dx \right) \phi_{i,\rho},$$

we have $1 \geq \lambda_{1,\rho} \geq \lambda_{2,\rho} \geq \cdots \geq 0$,

$$\int_0^1 \psi_{i,\rho}(x)dx = \int_0^1 \phi_{i,\rho}(y)dy = 0,$$

and

$$\int_0^1 \psi_{i,\rho}(x)\psi_{j,\rho}(x)dx = \int_0^1 \phi_{i,\rho}(y)\phi_{j,\rho}(y)dy = \delta_{i,j}.$$

Equivalently, we can write

$$f_\rho(x, y) = 1 + \sum_{i=1}^{\infty} \lambda_{i,\rho} \psi_{i,\rho}(x)\phi_{i,\rho}(y). \tag{2}$$

Thus,

$$A(\rho) = 1 + \sum_{i=1}^{\infty} \lambda_{i,\rho}^2 < \infty. \tag{3}$$

Define the following condition:

(HS) $A(\rho) < \infty$ where $\lambda_{1,\rho}$ is strictly less than 1.

Theorem 1. *Assume that the condition (HS) holds for $\rho = \rho_1, \rho_2$. Then there does not exist a consistent discrimination rule for ρ_1 versus ρ_2 based on the broken sample.*

Remark 1. The condition (HS) is a not a stringent one, and is satisfied by the majority of the commonly used bivariate statistical models. However, it will be demonstrated in a number of examples below that the condition (HS) can be violated, and for each of those examples consistent discrimination rules do exist. Hence a natural question is whether the violation of the condition (HS) necessarily implies the existence of consistent discrimination rules. We conjecture that the answer is affirmative, but we have not been able to show that.

At the heart of the proof of Theorem 1 is the following result, which deserves prominent attention in its own right. Denote by $g_{n,\rho}(\boldsymbol{x}, \boldsymbol{y})$ the density of the broken sample, i.e.

$$g_{n,\rho}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n!} \sum_{\pi} \prod_{i=1}^{n} f_\rho(x_i, y_{\pi_i}),$$

where the summation is taken over all permutations π of $1, \ldots, n$. By assumption, $g_{n,\rho_o}(\boldsymbol{x}, \boldsymbol{y}) \equiv 1$. As a result, $g_{n,\rho}(\boldsymbol{x}, \boldsymbol{y})$ can also be viewed as a likelihood ratio.

Theorem 2. *Let the condition (HS) hold for some ρ. Then*

$$\lim_{n\to\infty} P_{\rho_o}(g_{n,\rho}(\boldsymbol{X}, \boldsymbol{Y}) \le x) = P(\xi \le x) \textit{ for all } x,$$

where

$$\xi = \prod_{k=1}^{\infty} (1 - \lambda_{k,\rho}^2)^{-1/2} \exp\left(-\frac{1}{2} \sum_{k=1}^{\infty} \frac{\lambda_{k,\rho}}{1 - \lambda_{k,\rho}} U_k^2 + \frac{1}{2} \sum_{k=1}^{\infty} \frac{\lambda_{k,\rho}}{1 + \lambda_{k,\rho}} V_k^2 \right),$$

with the U_i, V_i denoting iid. standard normal random variables.

Observe that $\log \xi$ is a constant plus a weighted average of independent χ^2 random variables.

To give some insight into the conclusion of Theorem 1, we present the following perspective. Define

$$f_{\rho,\delta} = \delta \wedge f_\rho, \ \delta > 0,$$

Theorem 3. *Suppose that the condition (HS) holds for some ρ, and that each $\delta > 0$, $f_{\rho,\delta}$ is square integrable in the Riemann sense on $[0, 1] \times [0, 1]$. Then $\mathcal{P}_\rho \equiv \mathcal{P}_{\rho_o}$ (see section 1 for notation). Thus, the class of probability distributions $\mathcal{P}_\rho$ for which F_ρ satisfy these conditions are mutually equivalent.*

It is well-known that the probability distributions of any two Gaussian processes with the same sample space are equivalent if and only of the Kulback-Leibler information between the two is finite (cf. Hajék (1958)). Our proof therefore is based on the derivation of the Kulback-Leibler information between $\mathcal{P}_\rho$ and $\mathcal{P}_{\rho_o}$ in terms of f_ρ, where we show under the conditions stated in Theorem 3 that the Kulback-Leibler information between $\mathcal{P}_\rho$ and $\mathcal{P}_{\rho_o}$ is equal to

$$\sum_{i=1}^{\infty} \frac{\lambda_{i,\rho}^2}{1 - \lambda_{i,\rho}^2} < \infty.$$

The proofs of Theorems 1-3 are collected in section 3. We now present a few examples for both cases for which consistent estimation is possible and is not possible.

Example A. First we revisit the setting of DeGroot and Goel (1980). Let (U, V) have the bivariate normal distribution with standard marginals and correlation ρ and denote by ϕ_ρ the joint pdf. It is well known (see Cramér, 1946) that

$$\phi_\rho(u, v) = \sum_{k=0}^{\infty} \frac{\rho^k}{k!} H_k(u)\phi(u)H_k(v)\phi(v),$$

where ϕ is the standard normal pdf and $H_k(u) = (-1)^k e^{u^2/2} \frac{d^k}{du^k} e^{-u^2/2}$ is the k-th Hermite polynomial. Let f_ρ be the pdf of $(\Phi(U), \Phi(V))$. Then

$$f_\rho(x, y) = \sum_{k=0}^{\infty} \frac{\rho^k}{k!} H_k(Q(x))H_k(Q(y))$$

where Φ and Q are the standard normal cdf and quantile function, respectively. It is easy to check that (HS) holds for each ρ where $\lambda_{i,\rho} = |\rho|^i$. Thus, the question posed by DeGroot and Goel (1980) is completely answered.

Example B. Suppose that $y(x)$ is a monotone function such that $P_\rho(Y_1 = y(X_1)) = c(\rho)$. Then

$$\hat{c}(\rho) := n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} I(Y_j = y(X_i))$$

is obviously $\sqrt{n}$-consistent for $c(\rho)$. In this case, of course, (X_1, Y_1) does not have a joint density. One such example (cf. Chan and Loh, 2001) is to let J_i be iid. with $P(J_1 = 1) = 1 - P(J_1 = 0) = \rho$ where $\rho \in [0, 1]$ and

$$X_i = J_i U_i + (1 - J_i)V_i, \;\; Y_i = J_i U_i + (1 - J_i)W_i$$

where $U_i, V_i, W_i, 1 \le i \le n$ are iid; in this case, $P_\rho(Y = X) = \rho$.

Example C. Let

$$X_i = \rho U_i + (1-\rho)V_i,\ Y_i = \rho U_i + (1-\rho)W_i \tag{4}$$

where $\rho \in (0, 1)$ and the U_i, V_i, W_i are iid. standard Cauchy. In this case $A(\rho) = \infty$ and ρ can be consistently estimated. The intuition here is that when a large value of X is observed, the probability that it is due to a large U is ρ and the probability that it is due to a large V is $1-\rho$. Thus, the probability of finding a matching Y for a large X is roughly ρ. Indeed the following can be proved.

Theorem 4. *Let (X_i, Y_i) be defined by (4) and k_n and ε_n be positive constants such that $k_n \to \infty$, $k_n\varepsilon_n \to 0$ and $n\varepsilon_n/k_n \to \infty$. Then $A(\rho) = \infty$ for all $\rho \in (0, 1)$ and*

$$\hat{\rho}_n := \frac{1}{k_n}\sum_{i=1}^{k_n}\sum_{j=1}^{n} I(Y_j/X_{(i)} \in (1-\varepsilon_n, 1+\varepsilon_n)) \xrightarrow{p} \rho$$

where $X_{(i)}$ is the i-th largest value of $X_1, \ldots, X_n$.

The proof of Theorem 4 is given in section 3.4. This example can be easily extended to other heavy-tailed scenarios (cf. Resnick, 1987).

Example D. For $\rho \in [0, 1]$, define the density

$$f_\rho(x, y) = \rho^{-1} I(0 < x, y \le \rho) + (1-\rho)^{-1} I(\rho < x, y < 1),\ x, y \in (0, 1).$$

In this case, $\rho = 0, 1$ correspond to independence and $\rho = 1/2$ to maximal dependence. Let

$$g(x) = \sqrt{\frac{1-\rho}{\rho}}\, I(0 < x \le \rho) - \sqrt{\frac{\rho}{1-\rho}}\, I(\rho < x < 1).$$

Then $\int_0^1 g(x)dx = 0$, $\int_0^1 g^2(x)dx = 1$ and $\int_0^1\int_0^1 g(x) f_\rho(x, y) g(y) dx dy = 1$ so that $\lambda_{1,\rho} = 1$. Consistent discrimination between any two distinct values ρ_1, ρ_2 is trivial; an obvious such rule is $T_n = \rho_1$ if $\sum_{i=1}^n I_{\{X_i \le \rho_1\}} = \sum_{i=1}^n I_{\{Y_i \le \rho_1\}}$ and $T_n = \rho_2$ otherwise. However, it is not clear whether a consistent estimator exists.

3. Proofs

We will prove Theorems 2, 1, 3, and 4 in the subsections 3.1, 3.2, 3.3, and 3.4, respectively. For simplicity of notation, where no confusion is likely, we will sometimes suppress the reference to ρ in $\lambda_{i,\rho}$, $\psi_{i,\rho}$ and $\phi_{i,\rho}$.

3.1. Proof of Theorem 2

We need the following lemma.

Lemma 5. *Assume that the condition (HS) holds for some ρ. Then*

$$\lim_{n\to\infty} E_{\rho_o} g_{n,\rho}^2(\boldsymbol{X},\boldsymbol{Y}) = \prod_{k=1}^{\infty}(1-\lambda_{k,\rho}^2)^{-1} < \infty.$$

Proof. Clearly,

$$E_{\rho_o} g_{n,\rho}(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{n!}\sum_{\pi}\int\int\prod_{i=1}^{n} f_\rho(x_i, y_{\pi_i})dx_i dy_i = 1. \tag{5}$$

Also,

$$\begin{aligned} E_{\rho_o} g_{n,\rho}^2(\boldsymbol{X},\boldsymbol{Y}) &= \frac{1}{(n!)^2}\sum_{\pi,\pi'}\int\cdots\int\prod_{i=1}^{n} f_\rho(x_i, y_{\pi_i}) f_\rho(x_i, y_{\pi_i'})dx_i dy_i \\ &= \frac{1}{(n!)^2}\sum_{\pi,\pi'}\int\cdots\int\prod_{i=1}^{n}\sum_{k=0}^{\infty}\lambda_k^2\phi_k(y_{\pi_i})\phi_k(y_{\pi_i'})dy_i \\ &= \frac{1}{n!}\sum_{\pi}\int\cdots\int\prod_{i=1}^{n}\sum_{k=0}^{\infty}\lambda_k^2\phi_k(y_i)\phi_k(y_{\pi_i})dy_i. \end{aligned}$$

It is easy to verify that for a given permutation π

$$\int\cdots\int\prod_{i=1}^{n}\sum_{k=0}^{\infty}\lambda_k^2\phi_k(y_i)\phi_k(y_{\pi_i})dy_i = \prod_{t=1}^{\ell}\sum_{k=0}^{\infty}\lambda_k^{2i_t},$$

if the permutation π consists of ℓ cycles of sizes $i_1, \cdots, i_\ell$ ($\sum i_t = n$). Therefore, if j_t denotes the number of t among $\{i_1, \cdots, i_\ell\}$, then

$$E_{\rho_o} g_{n,\rho}^2(\boldsymbol{X},\boldsymbol{Y}) = \sum_{\ell=1}^{n}\sum_{\substack{j_1+j_2\cdots+j_n=\ell \\ j_1+2j_2\cdots+nj_n=n}}\prod_{t=1}^{n}\frac{1}{j_t!}\Big(\frac{1}{t}\sum_{k=0}^{\infty}\lambda_k^{2t}\Big)^{j_t} \tag{6}$$

In fact, it is easy to see from this that that $E_{\rho_o} g_{n,\rho}^2(\boldsymbol{X},\boldsymbol{Y})$ is the coefficient of z^n in the Taylor expansion of the function $\prod_{k=0}^{\infty}(1-z\lambda_k^2)^{-1}$. Choose $r \in (1, \lambda_1^{-2})$ and consider the Cauchy integral

$$\frac{1}{2\pi i}\oint_{|z|=r}\frac{1}{z^{n+1}}\prod_{k=0}^{\infty}(1-z\lambda_k^2)^{-1}dz$$

whose absolute value is less than

$$r^{-n}(r-1)^{-1}\prod_{k=1}^{\infty}(1-r\lambda_k^2)^{-1} \to 0.$$

By the Cauchy integral theorem, we conclude that the integral equals

$$E_{\rho_o} g^2_{n,\rho}(\boldsymbol{X}, \boldsymbol{Y}) - \prod_{k=1}^{\infty}(1-\lambda_k^2)^{-1}.$$

Hence, $E_{\rho_o} g^2_{n,\rho}(\boldsymbol{X}, \boldsymbol{Y}) \to \prod_{k=1}^{\infty}(1-\lambda_k^2)^{-1}$. □

Proof of Theorem 2. Suppose that $X_i, Y_i, i = 1, \cdots, n$ are $2n$ iid. random variables uniformly distributed over (0, 1). Write $W(x, y) = f_\rho(x, y) - 1 = \sum_{k=1}^{\infty} \lambda_k \psi_k(x)\phi_k(y)$. Then we have

$$\frac{1}{n!}\sum_{\pi}\prod_{i=1}^{n} f_\rho(X_i, Y_{\pi_i}) = \frac{1}{n!}\sum_{\pi}\prod_{i=1}(1 + W(X_i, Y_{\pi_i})) = 1 + \sum_{t=1}^{n} Q_t,$$

where

$$Q_t = \frac{1}{n!}\sum_{\pi}\sum_{1\le i_1<\cdots<i_t\le n}\prod_{j=1}^{t} W(X_{i_j}, Y_{\pi_{i_j}}).$$

We first simplify Q_t. Suppose $(s_1, \cdots, s_t)$ is a given ordered subset of $\{1, \cdots, n\}$. Since there are $(n-t)!$ different permutations such that $(\pi_{i_1}, \cdots, \pi_{i_t}) = (s_1, \cdots, s_t)$,

$$Q_t = \frac{(n-t)!}{n!}\sum_{1\le i_1<\cdots<i_t\le n}\sum_{s_1,\cdots,s_t}\prod_{j=1}^{t} W(X_{i_j}, Y_{s_j}).$$

It is easy to verify that all Q_t are of mean 0 and uncorrelated with each other, and

$$EQ_t^2 = \frac{1}{t!}\sum_{s_1,\cdots,s_t} E\prod_{j=1}^{t}\sum_{k=1}^{\infty}\lambda_k^2\phi_k(Y_j)\phi_k(Y_{s_j}).$$

Observe that the expectation in the above summation is 0 if $\{s_1, \cdots, s_t\} \neq \{1, \cdots, t\}$ due to the orthogonality of the functions ϕ_k's. Thus, for each non zero term, $\{s_1, \cdots, s_t\}$ can be considered as a permutation of $\{1, \cdots, t\}$. Classify the permutations by the numbers of cycles of sizes $j = 1, \cdots, t$. Suppose the number of cycles of size j is ν_j, then we have $\nu_1 + 2\nu_2 + \cdots + t\nu_t = t$ and

$$EQ_t^2 = \frac{1}{t!}\sum_{\nu_1+2\nu_2+\cdots+t\nu_t=t}\prod_{j=1}^{t}\frac{1}{\nu_j!}\left(\frac{1}{j}\sum_{k=1}^{\infty}\lambda_k^{2j}\right)^{\nu_j}.$$

Similar to the derivation of the limit of the left hand side of (6), we have

$$EQ_t^2 = \frac{1}{2\pi i t!}\oint_{|z|=r}\frac{1}{z^{t+1}}\prod_{k=1}^{\infty}(1 - z\lambda_k^2)^{-1}dz.$$

By choosing $r \in (1, \lambda_1^{-2})$, one sees that

$$EQ_t^2 \le \frac{1}{t!}r^{-t}\prod_{k=1}^{\infty}(1 - r\lambda_k^2)^{-1}.$$

Thus, it follows that

$$\sum_{t=T}^{n} Q_t \to 0 \text{ in } L_2 \tag{7}$$

for any (slow) $T = T_n \to \infty$. Thus, to prove the claim of the theorem, it suffices to deal with the joint distributional convergence of $(Q_1, \ldots, Q_t)$ for each fixed t.

By central limit theorem and the orthonormality of the ψ_k and ϕ_k, it follows that for any fixed positive integer m, we have

$$\left(n^{-1/2}\sum_{i=1}^{n}\psi_k(X_i), n^{-1/2}\sum_{i=1}^{n}\phi_k(Y_i), k=1,\cdots,m\right) \xrightarrow{d} N(0, I_{2m}) \tag{8}$$

In addition, by the Marcinkiewitz law of large numbers, with probability 1 we have

$$n^{-1}\sum_{i=1}^{n}\psi_{k_1}(X_i)\psi_{k_2}(X_i) \to \delta_{k_1,k_2},\ n^{-1}\sum_{i=1}^{n}\phi_{k_1}(Y_i)\phi_{k_2}(Y_i) \to \delta_{k_1,k_2}, \tag{9}$$

and for $s \geq 3$,

$$n^{-s/2}\sum_{i=1}^{n}\psi_{k_1}(X_i)\cdots\psi_{k_s}(X_i) \to 0,\ n^{-s/2}\sum_{i=1}^{n}\phi_{k_1}(Y_i)\cdots\phi_{k_s}(Y_i) \to 0. \tag{10}$$

Write $W_m(x, y) = \sum_{k=1}^{m} \lambda_k\psi_k(x)\phi_k(y)$ and define

$$\begin{aligned}
Q_{mt} &= \frac{(n-t)!}{n!}\sum_{1\leq i_1<\cdots<i_t\leq n}\ \sum_{s_1,\cdots,s_t}\prod_{j=1}^{t} W_m(X_{i_j}, Y_{s_j}) \\
&= \sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}\frac{(n-t)!}{n!}\sum_{1\leq i_1<\cdots<i_t\leq n}\ \sum_{s_1,\cdots,s_t}\prod_{j=1}^{t}\psi_{k_j}(X_{i_j})\phi_{k_j}(Y_{s_j}) \\
&= \sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}\frac{(n-t)!}{t!n!}\left(\sum_{i_1,\cdots,i_t}\prod_{j=1}^{t}\psi_{k_j}(X_{i_j})\right)\left(\sum_{s_1,\cdots,s_t}\prod_{j=1}^{t}\phi_{k_j}(Y_{s_j})\right).
\end{aligned} \tag{11}$$

Note that for fixed t, $(n-t)!/n! \sim n^{-t}$ in the above. To find the limit distribution of Q_{mt}, let us consider some special cases first. By (8), we have

$$Q_{m1} = \sum_{k=1}^{m}\lambda_k n^{-1}\sum_{i=1}^{n}\psi_k(X_i)\sum_{i=1}^{n}\phi_k(Y_i) \xrightarrow{d} \sum_{k=1}^{m}\lambda_k U_k V_k$$

and by (8) and (9)

$$\begin{aligned} Q_{m2} &= \sum_{k_1,k_2=1}^{m} \lambda_{k_1}\lambda_{k_2} \frac{1}{2n(n-1)} \Big[\Big(\sum_{i=1}^{n} \psi_{k_1}(X_i)\Big)\Big(\sum_{i=1}^{n} \psi_{k_2}(X_i)\Big) - \sum_{i=1}^{n} \psi_{k_1}(X_i))\psi_{k_2}(X_i)\Big] \\ &\quad \Big[\Big(\sum_{i=1}^{n} \phi_{k_1}(Y_i)\Big)\Big(\sum_{i=1}^{n} \phi_{k_2}(Y_i)\Big) - \sum_{i=1}^{n} \phi_{k_1}(Y_i))\phi_{k_2}(Y_i)\Big] \\ &\xrightarrow{d} \frac{1}{2} \sum_{k_1,k_2=1}^{m} \lambda_{k_1}\lambda_{k_2}(U_{k_1}U_{k_2} - \delta_{k_1,k_2})(V_{k_1}V_{k_2} - \delta_{k_1,k_2}) \\ &= \frac{1}{2}\left((\sum_{k=1}^{m} \lambda_k U_k V_k)^2 - \sum_{k=1}^{m} \lambda_k^2 (U_k^2 + V_k^2 - 1)\right). \end{aligned}$$

For $t = 3$, we have

$$\begin{aligned} &n^{-3/2} \sum_{i_1,i_2,i_3} \psi_{k_1}(X_{i_1})\psi_{k_2}(X_{i_2})\psi_{k_3}(X_{i_3}) \\ &= n^{-3/2} \prod_{j=1}^{3} \sum_{i=1}^{n} \psi_{k_j}(X_i) - n^{-3/2} \sum_{i=1}^{n} \psi_{k_1}(X_i)\psi_{k_2}(X_i) \sum_{i=1}^{n} \psi_{k_3}(X_i) \\ &\quad - n^{-3/2} \sum_{i=1}^{n} \psi_{k_1}(X_i)\psi_{k_3}(X_i) \sum_{i=1}^{n} \psi_{k_2}(X_i) - n^{-3/2} \sum_{i=1}^{n} \psi_{k_2}(X_i)\psi_{k_3}(X_i) \sum_{i=1}^{n} \psi_{k_1}(X_i) \\ &\quad + n^{-3/2} \sum_{i=1}^{n} \psi_{k_1}(X_i)\psi_{k_2}(X_i)\psi_{k_3}(X_i) \\ &\quad \xrightarrow{d} U_{k_1}U_{k_2}U_{k_3} - \delta_{k_1,k_2}U_{k_3} - \delta_{k_1,k_3}U_{k_2} - \delta_{k_2,k_3}U_{k_1}. \end{aligned}$$

Similarly,

$$\begin{aligned} &n^{-3/2} \sum_{s_1,s_2,s_3} \phi_{k_1}(Y_{s_1})\phi_{k_2}(Y_{s_2})\phi_{k_3}(Y_{s_3}) \\ &\xrightarrow{d} V_{k_1}V_{k_2}V_{k_3} - \delta_{k_1,k_2}V_{k_3} - \delta_{k_1,k_3}V_{k_2} - \delta_{k_2,k_3}V_{k_1}. \end{aligned}$$

Therefore,

$$\begin{aligned} Q_{m3} &\xrightarrow{d} \frac{1}{6}\left[\left(\sum_{k=1}^{m} \lambda_k U_k V_k\right)^3 - 3\left(\sum_{k=1}^{m} \lambda_k U_k V_k\right)\left(\sum_{k=1}^{m} \lambda_k^2 (U_k^2 + V_k^2 - 1)\right)\right. \\ &\quad \left. + 6 \sum_{k=1}^{m} \lambda_k^3 U_k V_k \right]. \end{aligned}$$

In general, by the inclusion-exclusion identity and noticing (10), we have for a given $\mathbf{k} = (k_1, \cdots, k_t)$,

$$
\begin{aligned}
& n^{-t/2} \sum_{i_1,\cdots,i_t} \prod_{j=1}^{t} \psi_{k_j}(X_{i_j}) \\
& \xrightarrow{d} \prod_{j=1}^{t} U_{k_j} - \sum_{\{j_1,j_2\}\subset\{1,\cdots,t\}} \delta_{k_{j_1},k_{j_2}} \prod_{j\notin\{j_1,j_2\}} U_{k_j} \\
& \quad + \sum_{\{j_1,j_2\}\cup\{j_3,j_4\}\subset\{1,\cdots,t\}} \delta_{k_{j_1},k_{j_2}} \delta_{k_{j_3},k_{j_4}} \prod_{j\notin\{j_1,\cdots,j_4\}} U_{k_j} \\
& \quad - \sum_{\{j_1,j_2\}\cup\{j_3,j_4\}\cup\{j_5,j_6\}\subset\{1,\cdots,t\}} \delta_{k_{j_1},k_{j_2}} \delta_{k_{j_3},k_{j_4}} \delta_{k_{j_5},k_{j_6}} \prod_{j\notin\{j_1,\cdots,j_6\}} U_{k_j} + \cdots \\
& \quad =: S_{t,\mathbf{k},1} - S_{t,\mathbf{k},2} + \cdots + (-1)^{[t/2]} S_{t,\mathbf{k},[t/2]+1},
\end{aligned}
$$

where the sum in $S_{t,\mathbf{k},1+\ell}, \ell \geq 1$, runs over all possible ℓ pairs of indices $\{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\}$, in which the 2ℓ indices are distinct, with the understanding that, for example, $\{\{j_1, j_2\}, \{j_3, j_4\}\}$ and $\{\{j_1, j_3\}, \{j_2, j_4\}\}$ are two different partitions. By symmetry, we have

$$
\begin{aligned}
& n^{-t/2} \sum_{s_1,\cdots,s_t} \prod_{j=1}^{t} \phi_{k_j}(Y_{s_j}) \\
& \xrightarrow{d} \prod_{j=1}^{t} V_{k_j} - \sum_{\{j_1,j_2\}\subset\{1,\cdots,t\}} \delta_{k_{j_1},k_{j_2}} \prod_{j\notin\{j_1,j_2\}} V_{k_j} \\
& \quad + \sum_{\{j_1,j_2\}\cup\{j_3,j_4\}\subset\{1,\cdots,t\}} \delta_{k_{j_1},k_{j_2}} \delta_{k_{j_3},k_{j_4}} \prod_{j\notin\{j_1,\cdots,j_4\}} V_{k_j} \\
& \quad - \sum_{\{j_1,j_2\}\cup\{j_3,j_4\}\cup\{j_5,j_6\}\subset\{1,\cdots,t\}} \delta_{k_{j_1},k_{j_2}} \delta_{k_{j_3},k_{j_4}} \delta_{k_{j_5},k_{j_6}} \prod_{j\notin\{j_1,\cdots,j_6\}} V_{k_j} + \cdots \\
& \quad =: T_{t,\mathbf{k},1} - T_{t,\mathbf{k},2} + \cdots + (-1)^{[t/2]} T_{t,\mathbf{k},[t/2]+1}.
\end{aligned}
$$

Substituting these two limits into (11), we get

$$
Q_{mt} \xrightarrow{d} \frac{1}{t!} \sum_{\ell=0}^{[t/2]} \sum_{\ell'=0}^{[t/2]} (-1)^{\ell+\ell'} \sum_{k_1,\cdots,k_t=1}^{m} \prod_{j=1}^{t} \lambda_{k_j} S_{t,\mathbf{k},\ell+1} T_{t,\mathbf{k},\ell'+1}. \tag{12}
$$

Then, letting $m \to \infty$, we get the limit distribution of Q_t.

We now proceed to simplify the limiting distribution of Q_t. First note that

$$\begin{aligned}
&\sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}S_{t,\mathbf{k},1}T_{t,\mathbf{k},1}\\
&\quad=\sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}\prod_{j=1}^{t}U_{k_j}\prod_{j=1}^{t}V_{k_j}\\
&\quad=\left(\sum_{k=1}^{m}\lambda_k U_k V_k\right)^t\to\left(\sum_{k=1}^{\infty}\lambda_k U_k V_k\right)^t\quad\text{as } m\to\infty. \qquad (13)
\end{aligned}$$

Next,

$$\begin{aligned}
&\sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}S_{t,\mathbf{k},1}T_{t,\mathbf{k},\ell+1}\\
&=\sum_{\cup_{s=1}^{\ell}\{j_{2s-1},j_{2s}\}\subset\{1,\cdots,t\}}\ \sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}U_{k_j}\prod_{s=1}^{\ell}\delta_{k_{2s-1},k_{2s}}\prod_{j\notin\{j_1,\cdots,j_{2s}\}}V_{k_j}\\
&=\frac{t!}{2^{\ell}\ell!(t-2\ell)!}\left(\sum_{k=1}^{m}\lambda_k^2 U_k\right)^{\ell}\left(\sum_{k=1}^{m}\lambda_k^2 U_k V_k\right)^{t-2\ell}\\
&\to\frac{t!}{2^{\ell}\ell!(t-2\ell)!}\left(\sum_{k=1}^{\infty}\lambda_k^2 U_k^2\right)^{\ell}\left(\sum_{k=1}^{\infty}\lambda_k U_k V_k\right)^{t-2\ell}. \qquad (14)
\end{aligned}$$

More generally, for $\ell,\ell'\geq 1$,

$$\begin{aligned}
&\sum_{k_1,\cdots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}S_{t,\mathbf{k},\ell+1}T_{t,\mathbf{k},\ell'+1}\\
&\quad=\sum_{\cup_{s=1}^{\ell}\{j_{2s-1},j_{2s}\}}\ \sum_{\cup_{s'=1}^{\ell'}\{i_{2s'-1},i_{2s'}\}}\ \sum_{k_1,\ldots,k_t=1}^{m}\prod_{j=1}^{t}\lambda_{k_j}\prod_{s=1}^{\ell}\delta_{k_{j_{2s-1}},k_{j_{2s}}}\prod_{s'=1}^{\ell'}\delta_{k_{i_{2s'-1}},k_{i_{2s'}}}\\
&\quad\prod_{j\notin\{j_1,\cdots,j_{2\ell}\}}U_{k_j}\prod_{i\notin\{i_1,\cdots,i_{2\ell'}\}}V_{k_i}. \qquad (15)
\end{aligned}$$

To classify various products in its expression, we first introduce some notation. Let $(u_1,u_2,\ldots,u_h)$ be a sequence of distinct integers from $\{1,2,\cdots,t\}$. It is said to be a

- a cycle of length h (even) if $\{\{v_1,v_2\},\cdots,\{v_{h-1},v_h\}\}\subset\{\{j_1,j_2\},\cdots,\{j_{2\ell-1},j_{2\ell}\}\}$ and $\{\{v_2,v_3\},\cdots,\{v_{h-2},v_{h-1}\},\{v_h,v_1\}\}\subset\{\{i_1,i_2\},\cdots,\{i_{2\ell-1},i_{2\ell}\}\}$, where $(v_1,v_2,\ldots,v_h)$ is any permutation of $(u_1,u_2,\ldots,u_h)$.
- a vv-chain of length h (even) if $\{\{u_1,u_2\},\cdots,\{u_{h-1},u_h\}\}\subset\{\{j_1,j_2\},\cdots,\{j_{2\ell-1},j_{2\ell}\}\}$ and $\{\{u_2,u_3\},\cdots,\{u_{h-2},u_{h-1}\}\}\subset\{\{i_1,i_2\},\cdots,\{i_{2\ell-1},i_{2\ell}\}\}$ and $u_1,u_h\notin\{i_1,\cdots,i_{2\ell'}\}$.

· a uu-chain of length h (even) if $\{\{u_2, u_3\}, \cdots, \{u_{h-2}, u_{h-1}\}\} \subset \{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\}$ and $\{\{u_1, u_2\}, \cdots, \{u_{h-1}, u_h\}\} \subset \{\{i_1, i_2\}, \cdots, \{i_{2\ell-1}, i_{2\ell}\}\}$ and $u_1, u_h \notin \{j_1, \cdots, j_{2\ell}\}$.
· a uv-chain of length h (odd) if $\{\{u_1, u_2\}, \cdots, \{u_{h-2}, u_{h-1}\}\} \subset \{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\}$ and $\{\{u_2, u_3\}, \cdots, \{u_{h-1}, u_h\}\} \subset \{\{i_1, i_2\}, \cdots, \{i_{2\ell-1}, i_{2\ell}\}\}$ and $u_1 \notin \{i_1, \cdots, i_{2\ell'}\}$, $u_h \notin \{j_1, \cdots, j_{2\ell}\}$, or alternatively.

An integer $u \le t$ is called a singleton if $u \notin \{j_1, \cdots, j_{2\ell}\} \cup \{i_1, \cdots, i_{2\ell'}\}$. A singleton corresponds to a factor $\sum_{k=1}^m \lambda_k U_k V_k$. A singleton can be considered as a uv-chain of length 1. In the sequel, we shall not specify singletons.

Observe that for each given partitions $\{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\}$ and $\{\{i_1, i_2\}, \cdots, \{i_{2\ell'-1}, i_{2\ell'}\}\}$, the set of numbers $\{1, 2, \ldots, t\}$ can be uniquely partitioned into disjoint sets each of which is a cycle or contains a uu, vv or a uv-chain. As a simple illustration, let $t = 12$ and consider the partitions

$$\{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\} = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}, \{10,11\}\},$$

and

$$\{\{i_1, i_2\}, \cdots, \{i_{2\ell'-1}, i_{2\ell'}\}\} = \{\{2,3\}, \{5,7\}, \{6,8\}, \{11,12\}\}.$$

Then $(5, 6, 7, 8)$ is a cycle, $(1, 2, 3, 4)$ is a vv-chain, 9 is a uv-chain (which is also a singleton), and $(10, 11, 12)$ is a uv-chain. Then it is easy to see that for this example,

$$\sum_{k_1,\ldots,k_t=1}^{m} \prod_{j=1}^{t} \lambda_{k_j} \prod_{s=1}^{\ell} \delta_{k_{j_{2s-1}}, k_{j_{2s}}} \prod_{s'=1}^{\ell'} \delta_{k_{i_{2s'-1}}, k_{i_{2s'}}} \prod_{j \notin \{j_1, \cdots, j_{2\ell}\}} U_{k_j} \prod_{i \notin \{i_1, \cdots, i_{2\ell'}\}} V_{k_i}$$
$$= \left(\sum_{k=1}^m \lambda_k^4\right) \left(\sum_{k=1}^m \lambda_k^4 V_k^2\right) \left(\sum_{k=1}^m \lambda_k U_k V_k\right) \left(\sum_{k=1}^m \lambda_k^3 U_k V_k\right).$$

Observe that each cycle, vv, uu and uv-chain of length h produce factors $\sum_{k=1}^m \lambda_k^h$, $\sum_{k=1}^m \lambda_k^h V_k^2$, $\sum_{k=1}^m \lambda_k^h U_k^2$ and $\sum_{k=1}^m \lambda_k^h U_k V_k$, respectively, in such a computation.

In general, for any given partitions $\{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\}$ and $\{\{i_1, i_2\}, \cdots, \{i_{2\ell'-1}, i_{2\ell'}\}\}$, denote by ν_{h1}, the numbers of cycles of length h, and ν_{h2}, nu_{h3} and ν_{h4} the number of vv, uu and uv-chains of length h, respectively. Note that $\sum_{h,i} h\nu_{hi} = t$. Then

$$\sum_{k_1,\ldots,k_t=1}^{m} \prod_{j=1}^{t} \lambda_{k_j} \prod_{s=1}^{\ell} \delta_{k_{j_{2s-1}}, k_{j_{2s}}} \prod_{s'=1}^{\ell'} \delta_{k_{i_{2s'-1}}, k_{i_{2s'}}} \prod_{j \notin \{j_1, \cdots, j_{2\ell}\}} U_{k_j} \prod_{i \notin \{i_1, \cdots, i_{2\ell'}\}} V_{k_i}$$
$$= \prod_{h \text{ even}} \left(\sum_{k=1}^m \lambda_k^h\right)^{\nu_{h1}} \left(\sum_{k=1}^m \lambda_k^h V_k^2\right)^{\nu_{h2}} \left(\sum_{k=1}^m \lambda_k^h U_k^2\right)^{\nu_{h3}} \prod_{h \text{ odd}} \left(\sum_{k=1}^m \lambda_k^h U_k V_k\right)^{\nu_{h4}}.$$

In addition to $\sum_{h,i} h\nu_{hi} = t$, we should also have the constraints $\nu_{h1} = \nu_{h2} = \nu_{h3} = 0$ for odd h, $\nu_{h4} = 0$ for even h, $\sum_h (2\nu_{h3} + \nu_{h4}) = t - 2\ell$ and $\sum_h (2\nu_{h2} + \nu_{h4}) = t - 2\ell'$. These constraints come directly from the definition; for example, each uv-chain and vv-chain correspond to one and two elements in $1, \ldots, t$, respectively, that are not in $\{\{i_1, i_2\}, \cdots, \{i_{2\ell'-1}, i_{2\ell'}\}\}$.

While each j-i combination, namely, $\{\{j_{2s-1}, j_{2s}\}, s = 1, \cdots, \ell\}$ and $\{\{i_{2s-1}, i_{2s}\}, s = 1, \cdots, \ell'\}$, generates a unique partition of $1, \ldots, t$, the converse is not true. For our proof it is easier to start from the partitions of $1, \ldots, t$ and compute the number of j-i combinations that correspond to the partition. Given a set of $\nu_{hi},\ h = 1, \cdots, t,\ i = 1, 2, 3, 4$, partition $\{1, 2, \cdots, t\}$ into groups $G_{hi}^{(s)}$ of sizes h, $s = 1, 2, \cdots, \nu_{hi},\ h = 1, \cdots, t,\ i = 1, 2, 3, 4$. The number of such partitions is

$$\frac{t!}{\prod_{h=1}^{t}\prod_{i=1}^{4} \nu_{hi}!(h!)^{\nu_{hi}}}.$$

For a given partition of $G_{hi}^{(s)}$, we compute the number of possible j-i pairs, namely, sets $\{\{j_{2s-1}, j_{2s}\}, s = 1, \cdots, w\}$ and $\{\{i_{2s-1}, i_{2s}\}, s = 1, \cdots, w'\}$, using elements of $G_{hi}^{(s)}$.

First consider any $G_{h1}^{(s)}$ with $h \geq 2$ even. Observe that there is a one-one correspondence between the circular permutations of the elements of $G_{h1}^{(s)}$ and the j-i pairs. Hence there correspond $(h-1)!$ possible j-i pairs.

Next consider any $G_{h2}^{(s)}$ with $h \geq 2$ even. For each permutation $(u_1, \cdots, u_h)$ of elements of $G_{h1}^{(s)}$, we construct the i-j pairs by putting $\{u_1, u_2\}, \cdots, \{u_{h-1}, u_h\}$ into the j-pairs, $\{u_2, u_3\}, \cdots, \{u_{h-2}, u_{h-1}\}$ into the i-pairs, and leaving u_1, u_h as the unpaired i's. Since the reverse permutation gives the same j-i partition, we have $h!/2$ ways to construct all possible j-i pairs.

Symmetrically, for each $G_{h3}^{(s)}$ with $h \geq 2$ even, we have $h!/2$ ways to construct all possible j-i pairs.

Finally consider any $G_{h4}^{(s)}$ with $h \geq 2$ odd. For each permutation $(u_1, \cdots, u_h)$ of elements of $G_{h1}^{(s)}$, we construct the i-j pairs by putting $\{u_1, u_2\}, \cdots, \{u_{h-2}, u_{h-1}\}$ into the j-pairs and $\{u_2, u_3\}, \cdots, \{u_{h-1}, u_h\}$ into the i-pairs and leaving u_1 as an unpaired i and u_h as an unpaired j. Thus, there are $h!$ ways to construct the possible $j - i$ pairs.

Put all integers in $G_{13}^{(s)}$ as both unpaired j's as well as the unpaired i's. Combining these, by (15), we have

$$\begin{aligned}
&\sum_{k_1,\cdots,k_t=1}^{m} \prod_{j=1}^{t} \lambda_{k_j} S_{t,\mathbf{k},\ell+1} T_{t,\mathbf{k},\ell'+1} \\
&= \sum{}^{*} t! \prod_{h \text{ even}} \frac{1}{\nu_{h1}!\nu_{h2}!\nu_{h3}!} \left(\frac{1}{h}\sum_{k=1}^{m}\lambda_k^h\right)^{\nu_{h1}} \left(\frac{1}{2}\sum_{k=1}^{m}\lambda_k^h V_k^2\right)^{\nu_{h2}} \\
&\quad \left(\frac{1}{2}\sum_{k=1}^{m}\lambda_k^h U_k^2\right)^{\nu_{h3}} \prod_{h \text{ odd}} \frac{1}{\nu_{h4}!} \left(\sum_{k=1}^{m}\lambda_k^h U_k V_k\right)^{\nu_{h4}} \\
&\to \sum{}^{*} t! \prod_{h \text{ even}} \frac{1}{\nu_{h1}!\nu_{h2}!\nu_{h3}!} \left(\frac{1}{h}\sum_{k=1}^{\infty}\lambda_k^h\right)^{\nu_{h1}} \left(\frac{1}{2}\sum_{k=1}^{\infty}\lambda_k^h V_k^2\right)^{\nu_{h2}}
\end{aligned}$$

$$\left(\frac{1}{2}\sum_{k=1}^{\infty}\lambda_k^h U_k^2\right)^{\nu_{h3}} \prod_{h \text{ odd}} \frac{1}{\nu_{h4}!}\left(\sum_{k=1}^{\infty}\lambda_k^h U_k V_k\right)^{\nu_{h4}},$$

where $\sum^*$ runs over all possible integers ν_{hi} subject to $\sum_{h,i} h\nu_{hi} = t$, $\nu_{h1} = \nu_{h2} = \nu_{h3} = 0$ for odd h, $\nu_{h4} = 0$ for even h, $\sum_h(2\nu_{h3} + \nu_{h4}) = t - 2\ell$ and $\sum_h(2\nu_{h2} + \nu_{h4}) = t - 2\ell'$.

Note that the results in (13) – (15) can also be written in the above form by letting $\ell = 0$ or $\ell' = 0$ or $\ell = \ell' = 0$. In addition, we have

$$(-1)^{\ell+\ell'} = (-1)^{t+\sum_{h=1}^{t}(\nu_{h2}+\nu_{h3}+\nu_{h4})}.$$

Thus,

$$Q_t \xrightarrow{d} (-1)^t \sum{}^{**} \prod_{h \text{ even}} \frac{1}{\nu_{h1}!\nu_{h2}!\nu_{h3}!}\left(\frac{1}{h}\sum_{k=1}^{\infty}\lambda_k^h\right)^{\nu_{h1}}\left(-\frac{1}{2}\sum_{k=1}^{\infty}\lambda_k^h V_k^2\right)^{\nu_{h2}}$$
$$\left(-\frac{1}{2}\sum_{k=1}^{\infty}\lambda_k^h U_k^2\right)^{\nu_{h3}} \prod_{h \text{ odd}} \frac{1}{\nu_{h4}!}\left(-\sum_{k=1}^{\infty}\lambda_k^h U_k V_k\right)^{\nu_{h4}},$$

where $\sum^{**}$ runs over all possible integers ν_{hi} subject to $\sum_{h,i} h\nu_{hi} = t$, $\nu_{h1} = \nu_{h2} = \nu_{h3} = 0$ for odd h, $\nu_{h4} = 0$ for even h. We further simplify it as

$$\begin{aligned} Q_t &\xrightarrow{d} (-1)^t \sum{}^{***} \prod_{h \text{ even}} \frac{1}{\nu_{h0}!}\left(\sum_{k=1}^{\infty}\lambda_k^h\Big(\frac{1}{h} - \frac{1}{2}U_k^2 - \frac{1}{2}V_k^2\Big)\right)^{\nu_{h0}} \\ &\qquad \prod_{h \text{ odd}} \frac{1}{\nu_{h4}!}\left(-\sum_{k=1}^{\infty}\lambda_k^h U_k V_k\right)^{\nu_{h4}} \\ &=: \ \xi_t, \end{aligned}$$

where the $\sum^{***}$ is taken for $\sum_h h(\nu_{h0} + \nu_{h4}) = t$.

Note that ξ_t is the coefficient of z^t in the Taylor expansion of

$$H(z) = \exp\left(\sum_{\substack{h\ge 2\\ \text{even}}}(-z)^h \sum_{k=1}^{\infty}\lambda_k^h\Big(\frac{1}{h} - \frac{1}{2}U_k^2 - \frac{1}{2}V_k^2\Big) - \sum_{\substack{h\ge 1\\ \text{odd}}}(-z)^h\sum_{k=1}^{\infty}\lambda_k^h U_k V_k\right).$$

Straightforward algebra shows that

$$\begin{aligned} \sum \xi_t &= H(1) = \prod_{k=1}^{\infty}(1-\lambda_{k,\rho}^2)^{-1/2} \\ &\qquad \exp\left(-\frac{1}{2}\sum_{k=1}^{\infty}\frac{\lambda_{k,\rho}^2}{1-\lambda_{k,\rho}^2}(U_k^2 + V_k^2) + \sum_{k=1}^{\infty}\frac{\lambda_{k,\rho}}{1-\lambda_{k,\rho}^2}U_k V_k\right). \end{aligned}$$

Making an orthogonal transformation with $U_k' = (U_k + V_k)/\sqrt{2}$, $V_k' = (U_k - V_k)/\sqrt{2}$, the right hand side is easily seen to have the same distribution as ξ described in the theorem. This concludes the proof. □

3.2. *Proof of Theorem 1*

Let $\mathcal{T}_n$ be a consistent discrimination rule for ρ_1 versus ρ_2. Write, for any fixed $M \in (0, \infty)$,

$$\begin{aligned} P_{\rho_1}(\mathcal{T}_n = \rho_1) = {} & P_{\rho_1}(\mathcal{T}_n = \rho_1, g_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) > M g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y})) \\ & + P_{\rho_1}(\mathcal{T}_n = \rho_1, g_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) \le M g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y})). \end{aligned}$$

Suppose that $P_{\rho_1}(g_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) > M g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y})) \not\to 1$. Since $P_{\rho_1}(\mathcal{T}_n = \rho_1) \to 1$, we have $P_{\rho_1}(\mathcal{T}_n = \rho_1, g_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) \le M g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y})) \not\to 0$. Thus

$$\begin{aligned} \limsup_{n\to\infty} P_{\rho_2}(\mathcal{T}_n = \rho_1) \ge {} & \limsup_{n\to\infty} M^{-1} P_{\rho_1}(\mathcal{T}_n = \rho_1, g_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) \\ & \le M g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y})) > 0. \end{aligned}$$

This contradicts the assumption that $\mathcal{T}_n$ is consistent. Hence, we must have

$$P_{\rho_1}(g_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) > M g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y})) \to 1 \text{ for all } M \in (0, \infty). \tag{16}$$

Now,

$$\begin{aligned} E_{\rho_o} g^2_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) & \ge \int g^2_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) I\big(g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) > M g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y})\big) d\boldsymbol{x} d\boldsymbol{y} \qquad (17) \\ & \ge M \int g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y}) I\big(g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) > M g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y})\big) d\boldsymbol{x} d\boldsymbol{y} \\ & \ge Md \int g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) I\big(g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) > M g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y}),\ g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y}) > d\big) d\boldsymbol{x} d\boldsymbol{y} \\ & \ge Md \int g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) I\big(g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) > M g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y})\big) d\boldsymbol{x} d\boldsymbol{y} \\ & \quad - Md \int g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) I\big(g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y}) \le d\big) d\boldsymbol{x} d\boldsymbol{y}. \end{aligned}$$

Choose $M = M_n \to \infty$ so slowly that (16) still holds when M is replaced by M_n. Then choose $d = d_n \to 0$ so slowly such that $M_n d_n \to \infty$. By Theorem 2 and the fact that ξ has a continuous distribution,

$$P_{\rho_o}(g_{n,\rho_2}(\boldsymbol{X}, \boldsymbol{Y}) \le d) \to 0.$$

Hence, by Lemma 5 and the Cauchy-Schwarz inequality,

$$\int g_{n,\rho_1}(\boldsymbol{x}, \boldsymbol{y}) I\big(g_{n,\rho_2}(\boldsymbol{x}, \boldsymbol{y}) \le d\big) d\boldsymbol{x} d\boldsymbol{y} \to 0. \tag{18}$$

It follows from (16), (18), and the choices of M, d that the last expression in (17) tends to ∞, which implies that

$$E_{\rho_o} g^2_{n,\rho_1}(\boldsymbol{X}, \boldsymbol{Y}) \to \infty.$$

This contradicts Lemma 5 and concludes the proof. □

3.3. Proof of Theorem 3

We begin by mentioning the following result due to Hajék (1958) specialized and simplified to our setting. See also Grenander (1981) and Rozanov (1971). Let $\mathcal{F}$ be the σ-field of the product space $D[0, 1]\times D[0, 1]$. Let Q_1 and Q_2 be two probability measures on $(D[0, 1] \times D[0, 1], \mathcal{F})$ which each correspond to the distribution of a Gaussian process. Let $\mathcal{F}_n$ be a sequence of sub σ-fields of $\mathcal{F}$ with $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n)$. The Kulback-Leibler information number between Q_1, Q_2 with respect to $\mathcal{F}_n$ is

$$J_n = E_{Q_1}(-\log q_n) + E_{Q_2}(\log q_n)$$

where q_n is the Radon-Nikodym derivative of the absolutely continuous part of Q_2 with respect to Q_1 on $(\Omega, \mathcal{F}_n)$.

Theorem 6. *Q_1 and Q_2 are equivalent on $(D[0, 1]\times D[0, 1], \mathcal{F})$ iff $\sup_n J_n < \infty$.*

Proof of Theorem 3. First we show that

$$\lim_{m\to\infty} \min_{1\le i\le m} \max_{1\le j\le m} m \int_{x=\frac{i-1}{m}}^{\frac{i}{m}} \int_{y=\frac{j-1}{m}}^{\frac{j}{m}} f_\rho(x, y)dxdy = 0. \tag{19}$$

To do that, choose δ_m and split the integral in (19) into two parts according to $f_\rho < \delta_m$ or not. Then

$$\begin{aligned} &\min_{1\le i\le m} \max_{1\le j\le m} m \int_{x=\frac{i-1}{m}}^{\frac{i}{m}} \int_{y=\frac{j-1}{m}}^{\frac{j}{m}} f_\rho(x, y)dxdy \\ &\le \frac{\delta_m}{m} + \frac{m}{\delta_m} \min_i \int_{(i-1)/m}^{i/m} \int_0^1 f_\rho^2(x, y)dxdy. \end{aligned} \tag{20}$$

By the condition (HS),

$$\min_i \int_{(i-1)/m}^{i/m} \int_0^1 f_\rho^2(x, y)dxdy \to 0.$$

Thus, we may select a sequence δ_m such that $\delta_m/m \to 0$ so slowly that the second term of (20) tends to 0, which proves (19). It follows from (19) that there exists a sequence $i_m \in \{1, \dots, m\}$ such that

$$\lim_{m\to\infty} \max_{1\le j\le m} mP\left(\frac{i_m - 1}{m} < X \le \frac{i_m}{m}, \frac{j-1}{m} < Y \le \frac{j}{m}\right) = 0. \tag{21}$$

We will apply Theorem 6 by letting $\mathcal{F}_m = \sigma\{Z_1(\frac{i-1}{m}) - Z_1(\frac{i}{m}), Z_2(\frac{i-1}{m}) - Z_2(\frac{i}{m}) : i \in \{1, \dots, m\} - \{i_m\}\}$ and $\mathcal{F} = \sigma\{Z_1(t), Z_2(t), t \in (0, 1)\}$. Since almost all paths of the Gaussian process are continuous, we have $\mathcal{F} = \sigma(\cup \mathcal{F}_m)$. The joint pdf of

$$\begin{aligned} \boldsymbol{U} = \Bigg(&Z_1\left(\frac{i-1}{m}\right) - Z_1\left(\frac{i}{m}\right), Z_2\left(\frac{i-1}{m}\right) \\ &- Z_2\left(\frac{i}{m}\right) : i \in \{1, \dots, m\} - \{i_m\}\Bigg) \end{aligned}$$

is $(2\pi)^{-(m-1)/2}|\Sigma_{m,\rho}|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{u}'\Sigma_{m,\rho}^{-1}\boldsymbol{u}\right)$ where $\Sigma_{m,\rho}$ is the covariance matrix of $\boldsymbol{U}$, namely

$$\Sigma_{m,\rho} = \begin{bmatrix} A_m & B_{m,\rho} \\ B'_{m,\rho} & A_m \end{bmatrix}$$

where A_m $((m-1)\times(m-1))$ is the auto-covariance of $(Z_k(\frac{i-1}{m}) - Z_k(\frac{i}{m}), \in \{1,\dots,m\} - \{i_m\}, k = 1, 2)$, and $B_{m,\rho}$ $((m-1)\times(m-1))$ is the cross-covariance of $(Z_1(\frac{i-1}{m}) - Z_1(\frac{i}{m}), i \in \{1,\dots,m\} - \{i_m\})$ with $(Z_2(\frac{i-1}{m}) - Z_2(\frac{i}{m}) : i \in \{1,\dots,m\} - \{i_m\})$. Clearly, $A_m = \frac{1}{m}I_{m-1} - \frac{1}{m^2}\mathbf{1}\mathbf{1}'$ and

$$B_{m,\rho} = \left[\mathcal{P}_\rho\left(\frac{i-1}{m} < X \le \frac{i}{m}, \frac{j-1}{m} < Y \le \frac{j}{m}\right) - \frac{1}{m^2},\ i, j \in \{1,\dots,m\} - \{i_m\}\right], \tag{22}$$

where $\mathbf{l}$ the $(m-1)$-vector of all elements 1. It is well-known (cf. Parzen, 1963) that the Kulback-Leibler information between $\mathcal{P}_\rho$ and $\mathcal{P}_{\rho_o}$ with respect to $\mathcal{F}_m$ is

$$J_m = \frac{1}{2}\{\mathrm{tr}(\Sigma_{m,\rho_o}^{-1}\Sigma_{m,\rho}) + \mathrm{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho_o}) - 4(m-1)\}.$$

By assumption $B_{m,\rho_o} = 0$ and hence

$$\mathrm{tr}(\Sigma_{m,\rho_o}^{-1}\Sigma_{m,\rho}) = 2(m-1).$$

As a result,

$$J_m = \frac{1}{2}\mathrm{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho_o}) - (m-1)$$

which, by Lemma 8 below, is equal to

$$\begin{aligned}\mathrm{tr}\{[I_{m-1} - A_m^{-1}B_{m,\rho}A_m^{-1}B'_{m,\rho}]^{-1}\} - (m-1) &= \sum_{i=1}^{m-1}(1-\lambda_{m,i}^2)^{-1} - (m-1)\\ &= \sum_{i=1}^{m-1}\frac{\lambda_{m,i}^2}{1-\lambda_{m,i}^2},\end{aligned}$$

where $\lambda_{m,1} \ge \dots \ge \lambda_{m,m-1} \ge 0$ are the singular values of $A_m^{-1/2}B_{m,\rho}A_m^{-1/2}$. Note that these are the canonical correlations between $(Z_1(\frac{i-1}{m}) - Z_1(\frac{i}{m}), i \in \{1,\dots,m\} - \{i_m\})$ and $(Z_2(\frac{i-1}{m}) - Z_2(\frac{i}{m}) : i \in \{1,\dots,m\} - \{i_m\})$ and so $\lambda_{m,1} \le 1$. It follows from (24) of Lemma 7 that

$$J_m \to \sum_{i=1}^{\infty}\frac{\lambda_i^2}{1-\lambda_i^2} < \infty$$

where $\lambda_i = \lambda_{i,\rho}$. By Theorem 6, we have $\mathcal{P}_\rho \equiv \mathcal{P}_{\rho_o}$. This concludes the proof. □

Lemma 7. *Assume that the conditions of Theorem 3 hold. Let $\lambda_{m,i}$ be as defined in the proof of Theorem 3 and $\lambda_i = \lambda_{i,\rho}$. Then*

$$\lim_{m\to\infty}\sum_{i=1}^{\infty}(\lambda_{m,i}-\lambda_i)^2 = 0, \tag{23}$$

where $\lambda_{m,i} = 0$ if $i \ge m$, and

$$\lim_{m\to\infty}\sum_{i=1}^{m-1}\frac{\lambda_{m,i}^2}{1-\lambda_{m,i}^2} = \sum_{i=1}^{\infty}\frac{\lambda_i^2}{1-\lambda_i^2} < \infty. \tag{24}$$

Proof. We first prove (23). Let A_m and $B_{m,\rho}$ be as given in the proof of Theorem 3. Note that

$$A_m^{-1} = (m^{-1}(I_{m-1} - m^{-1}\mathbf{l}\mathbf{l}'))^{-1} = m(I_{m-1} + \mathbf{l}\mathbf{l}') = mM\begin{bmatrix} m & 0 \\ 0 & I_{m-2}\end{bmatrix}M',$$

where M is an $(m-1)\times(m-1)$ orthogonal matrix with the first column $\mathbf{l}/\sqrt{m-1}$. This implies that

$$A_m^{-1/2} = \sqrt{m}M\begin{bmatrix} \sqrt{m} & 0 \\ 0 & I_{m-2}\end{bmatrix}M',$$

Clearly, $\{\lambda_{m,i}\}$ are the singular values of

$$\begin{aligned} A_m^{-1/2}B_{m,\rho}A_m^{-1/2} &= mM\begin{bmatrix} \sqrt{m} & 0 \\ 0 & I_{m-2}\end{bmatrix}M'B_{m,\rho}M\begin{bmatrix} \sqrt{m} & 0 \\ 0 & I_{m-2}\end{bmatrix}M' \\ &= mM\begin{bmatrix} \frac{m}{m-1}\mathbf{l}'B_{m,\rho}\mathbf{l} & \sqrt{\frac{m}{m-1}}\mathbf{l}'B_{m,\rho}M_1 \\ \sqrt{\frac{m}{m-1}}M_1'B_{m,\rho}\mathbf{l} & M_1'B_{m,\rho}M_1\end{bmatrix}M', \end{aligned}$$

where M_1 consists of the last $(m-2)$ columns of M. Thus, $\lambda_{m,1},\cdots,\lambda_{m,m-1}$ are also the singular values of

$$m\begin{bmatrix} \frac{m}{m-1}\mathbf{l}'B_{m,\rho}\mathbf{l} & \sqrt{\frac{m}{m-1}}\mathbf{l}'B_{m,\rho}M_1 \\ \sqrt{\frac{m}{m-1}}M_1'B_{m,\rho}\mathbf{l} & M_1'B_{m,\rho}M_1\end{bmatrix}.$$

Let $\eta_{m,1} \ge \cdots \ge \eta_{m,m-1}$ be the singular values of $mM'B_{m,\rho}M$. By Lemma A1, we have

$$\begin{aligned} \sum_{i=1}^{m-1}(\lambda_{m,i}-\eta_{m,i})^2 &\le m^2\Big[(\mathbf{l}'B_{m,\rho}\mathbf{l})^2 + \frac{(\sqrt{m}-1)^2}{m-1}\mathbf{l}'B_{m,\rho}M_1M_1'B_{m,\rho}'\mathbf{l}\Big] \\ &\le m^2\Big[(\mathbf{l}'B_{m,\rho}\mathbf{l})^2 + \mathbf{l}'B_{m,\rho}B_{m,\rho}'\mathbf{l}\Big]. \end{aligned}$$

We shall show the right hand side of the above inequality tends to 0. By (21) and (22),

$$\begin{aligned}\mathbf{l}'B_{m,\rho}\mathbf{l} &= \operatorname{cov}\left(I\left(\frac{i_m-1}{m} < X \le \frac{i_m}{m}\right), I\left(\frac{i_m-1}{m} < Y \le \frac{i_m}{m}\right)\right)\\ &= P\left(\frac{i_m-1}{m} < X \le \frac{i_m}{m}, \frac{i_m-1}{m} < Y \le \frac{i_m}{m}\right) - m^{-2} = o(m^{-1}).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbf{l}'B_{m,\rho}B'_{m,\rho}\mathbf{l} &= \sum_{j\neq i_m} \operatorname{cov}^2\left(I\left(\frac{i_m-1}{m} < X \le \frac{i_m}{m}\right), I\left(\frac{j-1}{m} < Y \le \frac{j}{m}\right)\right)\\ &\le 2\sum_{j\neq i_m}\left[P^2\left(\frac{i_m-1}{m} < X \le \frac{i_m}{m}, \frac{j-1}{m} < Y \le \frac{j}{m}\right) + m^{-4}\right]\\ &\le 2o(m^{-1})\sum_{j\neq i_m} P\left(\frac{i_m-1}{m} < X \le \frac{i_m}{m}, \frac{j-1}{m} < Y \le \frac{j}{m}\right) + o(m^{-3})\\ &= o(m^{-2})\end{aligned}$$

where the second inequality follows again from (21). Thus, to complete the proof of the lemma, we need only show that

$$\sum_{i=1}^{\infty}(\eta_{m,i} - \lambda_i)^2 \to 0,$$

where $\eta_{m,i} = 0$ if $i \ge m$. Let the singular decomposition of $mB_{m,\rho}$ be given by

$$mB_{m,\rho} = \sum_{\substack{1\le i\le m\\ i\neq i_m}} \eta_{m,i}\boldsymbol{\psi}_{m,i}\boldsymbol{\phi}'_{m,i}, \tag{25}$$

where $\boldsymbol{\psi}_{m,i} = (\psi_{m,i,1}, i \in \{1,\ldots,m\}-\{i_m\})'$ and $\boldsymbol{\phi}_{m,i} = (\phi_{m,i,1}, i \in \{1,\ldots,m\}-\{i_m\})'$. Now, define

$$p_m(x,y) = \sum_{\substack{1\le i,j\le m\\ i,j\neq i_m}} (\xi_{i,j} - 1) I_{[\frac{i-1}{m},\frac{i}{m})\times[\frac{j-1}{m},\frac{j}{m})}(x,y), \tag{26}$$

where $\xi_{i,j} = m^2 P(\frac{i-1}{m} < X \le \frac{i}{m}, \frac{j-1}{m} < Y \le \frac{j}{m})$, and also define

$$\begin{aligned}\psi_{m,i}(x) &= \sqrt{m}\sum_{\substack{1\le k\le m\\ k\neq i_m}} \psi_{m,i,k} I_{[\frac{k-1}{m},\frac{k}{m})}(x),\\ \phi_{m,i}(y) &= \sqrt{m}\sum_{\substack{1\le k\le m\\ k\neq i_m}} \phi_{m,i,k} I_{[\frac{k-1}{m},\frac{k}{m})}(y).\end{aligned}$$

Note that

$$\int_0^1 \psi_{m,i}(x)dx = \int_0^1 \phi_{m,i}(y)dy = 0,$$
$$\int_0^1 \psi_{m,i}(x)\psi_{m,j}(x)dx = \int_0^1 \phi_{m,i}(y)\phi_{m,j}(y)dy = \delta_{i,j}.$$

With these, it is easy to verify that the singular-value decomposition (25) can be rewritten as the singular-value decomposition of the linear transformation $T_{p_m} : g \to \int p_m(\cdot, y)g(y)dy$ as

$$p_m(x, y) = \sum_{\substack{1\le i\le m\\ i\ne i_m}} \eta_{m,i}\psi_{m,i}(x)\phi_{m,i}(y)$$

For any $\delta > 0$, define

$$\xi_{i,j,\delta} = m^2 \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} f_{\rho,\delta}(x, y)dxdy.$$

Thus, by Lemma A1 and (26), we have

$$\begin{aligned}
&\sum_{i=1}^{\infty}(\eta_{m,i} - \lambda_i)^2\\
&\le \int_0^1 \int_0^1 (p_m(x, y) - f_\rho(x, y) + 1)^2 dxdy\\
&= \int_0^1 \int_0^1 I\left(x \text{ or } y \in \left[\frac{i_m - 1}{m}, \frac{i_m}{m}\right)\right)(f_\rho(x, y) - 1)^2 dxdy\\
&+ \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (\xi_{i,j} - f_\rho(x, y))^2 dxdy.
\end{aligned}$$

The first term on the right-hand side tends to 0 as m tends to ∞ since f_ρ is square integrable. By the triangle inequality, the second term is bounded by $3(B_{m,\delta,1} + B_{m,\delta,2} + B_{m,\delta,3})$ where

$$B_{m,\delta,1} = \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (\xi_{i,j,\delta} - f_{\rho,\delta}(x, y))^2 dxdy,$$

$$B_{m,\delta,2} = \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (\xi_{i,j} - \xi_{i,j,\delta})^2 dxdy$$

$$B_{m,\delta,3} = \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (f_\rho(x, y) - f_{\rho,\delta}(x, y))^2 dxdy.$$

For each fixed δ, $B_{m,\delta,1} \to 0$ as m tends to ∞ by Riemann integrability of $f^2_{\rho,\delta}$. Applying the Cauchy-Schwarz inequality,

$$
\begin{aligned}
B_{m,\delta,2} &= m^2 \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \left(\int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (f_\rho(x,y) - f_{\rho,\delta}(x,y))dxdy \right)^2 \\
&\le \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (f_\rho(x,y) - f_{\rho,\delta}(x,y))^2 dxdy = B_{m,\delta,3}.
\end{aligned}
$$

Now,

$$
\begin{aligned}
B_{m,\delta,3} &= \sum_{\substack{1\le i,j\le m\\ i,j\ne i_m}} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (f_\rho(x,y) - f_{\rho,\delta}(x,y))^2 dxdy \\
&\le \int_0^1 \int_0^1 (f_\rho(x,y) - f_{\rho,\delta}(x,y))^2 dxdy.
\end{aligned}
$$

Since $0 \le f_\rho(x,y) - f_{\rho,\delta}(x,y) \le f_\rho(x,y)$, by dominated convergence theorem we conclude that

$$
\lim_{\delta\to\infty} \limsup_{m\to\infty} B_{m,\delta,2} = \lim_{\delta\to\infty} \limsup_{m\to\infty} B_{m,\delta,3} = 0.
$$

This concluded the proof of (23).

We next prove (24). Since $\lim_{m\to\infty} \lambda_{m,i} \to \lambda_i$ for all i by (23) and $\lambda_1 < 1$, it follows that for any $\epsilon \in (0, 1-\lambda_1)$, we have $\lambda_{m,i} \le \lambda_1 + \epsilon < 1$ for all i and all large m. Thus it is straightforward to conclude from (23) and (3) that that

$$
\sum_{i=1}^{m-1} \frac{\lambda^2_{m,i}}{1-\lambda^2_{m,i}} \to \sum_{i=1}^{\infty} \frac{\lambda^2_i}{1-\lambda^2_i} \le (1-\lambda_1)^{-1}(A(\rho)-1) < \infty.
$$

This concludes the proof. □

Lemma 8. *Assume that the conditions of Theorem 3 hold. Then for all large m,* $A_m - B_{m,\rho}A_m^{-1}B'_{m,\rho}$ *is invertible and hence*

$$
\mathrm{tr}(\Sigma^{-1}_{m,\rho}\Sigma_{m,\rho_o}) = 2\mathrm{tr}\{[I_{n(m)} - A_m^{-1}B_{m,\rho}A_m^{-1}B'_{m,\rho}]^{-1}\}.
$$

Proof. The first assertion follows simply from the fact that $\lambda_{m,1} \to \lambda_1 < 1$ by Lemma 7 and the condition (HS). To show the second assertion, for simplicity, we'll drop the indices m and ρ in A_m and $B_{m,\rho}$. By Theorem 8.5.11 of Harvill (1997),

$$
\begin{aligned}
\Sigma^{-1}_{m,\rho} &= \begin{bmatrix} A & B \\ B' & A \end{bmatrix}^{-1} \\
&= \begin{bmatrix} A^{-1} + A^{-1}B(A-B'A^{-1}B)^{-1}B'A^{-1} & -A^{-1}B(A-B'A^{-1}B)^{-1} \\ -(A-B'A^{-1}B)^{-1}B'A^{-1} & (A-B'A^{-1}B)^{-1} \end{bmatrix}.
\end{aligned}
$$

Hence

$$\begin{aligned}\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho_o} &= \begin{bmatrix} A & B \\ B' & A \end{bmatrix}^{-1} \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix} \\ &= \begin{bmatrix} I_n + A^{-1}B(A - B'A^{-1}B)^{-1}B' & -A^{-1}B(A - B'A^{-1}B)^{-1}A \\ -(A - B'A^{-1}B)^{-1}B' & (A - B'A^{-1}B)^{-1}A \end{bmatrix}.\end{aligned}$$

Thus,

$$\begin{aligned}&\mathrm{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho_o}) \\ &\quad = \mathrm{tr}(I_n + A^{-1}B(A - B'A^{-1}B)^{-1}B' + (A - B'A^{-1}B)^{-1}A) \\ &\quad = \mathrm{tr}(I_n) + \mathrm{tr}[(A - B'A^{-1}B)^{-1}(B'A^{-1}B - A)] + 2\mathrm{tr}[(A - B'A^{-1}B)^{-1}A] \\ &\quad = 2\mathrm{tr}[(A - B'A^{-1}B)^{-1}A] \\ &\quad = 2\mathrm{tr}[(I_n - A^{-1}BA^{-1}B')^{-1}].\end{aligned}$$

□

3.4. Proof of Theorem 4

First,

$$A(\rho) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_\rho^2(x, y)/(h(x)h(y))dxdy$$

where h is standard Cauchy pdf and

$$w_\rho(x, y) = \int_{-\infty}^{\infty} h(u)h\left(\frac{x - \rho u}{1 - \rho}\right)h\left(\frac{y - \rho u}{1 - \rho}\right)du$$

For $(x, y) \in [j, j + 1]^2$, $j \neq 0$,

$$w_\rho(x, y) \geq c\int_{|\rho u - j| < 1} h(u)du \geq c/j^2.$$

Hence,

$$A(\rho) \geq c\sum_{j\neq 0}\int_j^{j+1}\int_j^{j+1} \frac{1}{j^4 h(x)h(y)}dxdy = \infty.$$

Next we prove the consistency of the estimator $\hat{\rho}_n$. For convenience, drop n in k_n and ε_n. Let $i*$ be the index of $X_{(i)}$. Write

$$\hat{\rho}_n := S_n + R_n$$

where

$$S_n = \frac{1}{k}\sum_{i=1}^{k} I(Y_{i*}/X_{(i)} \in (1 - \varepsilon, 1 + \varepsilon))$$

and

$$R_n = \frac{1}{k}\sum_{i=1}^{k}\sum_{1\le j\le n, j\ne i*} I(Y_j/X_{(i)} \in (1-\varepsilon, 1+\varepsilon)).$$

Using the fact that, given $X_{(k+1)} = z$, $X_{(1)}, \dots, X_{(k)}$ are distributed as the order statistics of an iid. sample with pdf $h(x)I(x > z)/\int_z^\infty h(u)du$, we have

$$E(R_n|X_{(k+1)}) = n\int_{X_{(k+1)}}^{\infty} h(x)dx \int_{x(1-\varepsilon)}^{x(1+\varepsilon)} h(y)dy/\int_{X_{(k+1)}}^{\infty} h(u)du$$

It is therefore easy to see that

$$\begin{aligned} E[E(R_n|X_{(k+1)})I(X_{(k+1)} > 1)] &\le cn\varepsilon E[I(X_{(k+1)} > 1)/X_{(k+1)}] \\ &\le cn\varepsilon k/n \to 0. \end{aligned}$$

Since $I(X_{(k+1)} \le 1) \xrightarrow{p} 0$, we conclude that $R_n \xrightarrow{p} 0$. Next we will show that $S_n \xrightarrow{p} \rho$. Using the fact that, given $X_{(k+1)} = z$, $(X_{(1)}, Y_{(1)}), \dots, (X_{(k)}, Y_{(k)})$ are distributed as iid. with pdf $w_\rho(x, y)I(x > z)/\int_z^\infty h(u)du$, we obtain

$$\begin{aligned} E(S_n|X_{(k+1)}) &= \int_{X_{(k+1)}}^{\infty}\int_{x(1-\varepsilon)}^{x(1+\varepsilon)} w_\rho(x, y)dxdy/\int_{X_{(k+1)}}^{\infty} h(x)dx \\ &= \int_{X_{(k+1)}}^{\infty} h(x)P(Y \in [x(1-\varepsilon), x(1+\varepsilon)]|X \\ &= x)dx/\int_{X_{(k+1)}}^{\infty} h(x)dx. \end{aligned}$$

Let u_n be constants such that $\varepsilon u_n \to \infty$ and $u_n = o(n/k)$ so that $P(X_{(k+1)} < u_n) \to 0$. It is easy to show (cf. Resnick, 1987) that

$$\lim_{x\to\infty} P(Y_1/X_1 \in (1-\delta(x), 1+\delta(x))|X_1 = x) = \rho.$$

for any $\delta(x)$ with satisfying $\delta(x) \to 0$ and $x\delta(x) \to \infty$. Hence, for $X_{(k+1)} > u_n$ we have

$$E(S_n|X_{(k+1)}) \sim \rho$$

and, similarly,

$$\mathrm{var}(S_n|X_{(k)}) \le (1/k)E(S_n|X_{(k+1)}) \sim \rho/k.$$

It is then straightforward to conclude from these that $S_n \xrightarrow{p} \rho$. □

Appendix

The following technical result was applied in the proof of Lemma 7. The proof can be found in Lemma 2.7 of Bai (1999).

Lemma A1. (i) *Let A and B be $m \times n$ matrices with singular values λ_i and η_i (both in descending order) respectively. Then,*

$$\sum_{i=1}^{m\wedge n}(\lambda_i - \eta_i)^2 \leq tr[(A-B)(A-B)'].$$

(ii) *Let $\phi(s,t)$ and $\psi(s,t)$ be square integrable functions on* $[0,1]\times[0,1]$ *and let T_ϕ and T_ψ be two linear operators from $L^2[0,1]$ into itself defined by*

$$T_\phi(g) = \int_0^1 \phi(\cdot, y)g(y)dy \text{ and } T_\psi(g) = \int_0^1 \psi(\cdot, y)g(y)dy.$$

Let the λ_i and η_i be the singular values (both in descending order) of T_ϕ and T_ψ, respectively. Then,

$$\sum_{i=1}^{\infty}(\lambda_i - \eta_i)^2 \leq \int_0^1\int_0^1 [\phi(s,t) - \psi(s,t)]^2 ds dt.$$

□

References

Bai, Z.D.: Methodologies in spectral analysis of large dimensional random matrices. A rev. Statistica Sinica **9**, 611–677 (1999)

Chan, H.-P., Loh, W.-L.: A file linkage problem of DeGroot and Goel revisited. Statistica Sinica **11**, 1031–1045 (2001)

Cramér, H.: Mathematical methods of statistics. Princeton University Press, 1946

DeGroot, M.H., Goel, P.K.: Estimation of the correlation coefficient from a broken sample. Ann. Statist. **8**, 264–278 (1980)

Feller, W.: An Introduction to Probability and Its Applications. vol 2. Wiley, 1971

Grenander, U.: Abstract Inference. Wiley, 1981

Hajék, J.: On a property of normal distribution of any stochastic process (in Russian). Czechoslovak Math. J. **8**(83), 610–618 (1958), (An English translation appeared in American Mathematical Society Translations in Probability and Statistics, 1961)

Harvill, D.A.: Matrix Algebra from a Statistician's Perspective. Springer, 1997

Parzen, E.: Probability density functionals and reproducing kernel Hilbert spaces. M. Rosenblatt (ed.), Proceedings of the Symposium on Time Series Analysis, Wiley, 1963, pp. 155–169

Resnick, S.: Extreme Values, Regular Variation, and Point Processes. Springer, 1987

Riesz, F., Sz.-Nagy, B.: Functional Analysis. Translated from the 2d French ed. by Leo F. Boron. Ungar, 1955

Rozanov, J.A.: Infinite Dimensional Gaussian Distributions. English translation published by American Mathematical Society, 1971