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 – © Springer-Verlag 2004

Abstract. Suppose that $(X_i, Y_i), i = 1, 2, \ldots, n$, are iid. random vectors with uniform marginals and a certain joint distribution $F_\rho$, where $\rho$ 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 $\rho$ be consistently estimated? This is an extension of a problem considered in DeGroot and Goel (1980) which focused on the bivariate normal distribution with $\rho$ being the correlation. In this paper we show that consistent discrimination between two distinct parameter values $\rho_1$ and $\rho_2$ is impossible if the density $f_\rho$ of $F_\rho$ 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 $\rho_2$. 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 $\rho$ and let $F_\rho$ be the joint cdf. One can think of $\rho$ as a measure of association such as the correlation. We assume that the parameter space contains a specific value $\rho_o$, which corresponds to the independence of the marginals. Let $(X_i, Y_i), i = 1, 2, \ldots, n$, 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 $F_\rho$, 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
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 \( \rho \) 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 \(\rho_0\) 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 \(\rho_1\) and \(\rho_2\) 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

\[
\lim_{n \to \infty} P_{\rho_1}(T_n = \rho_1) = 1, \quad i = 1, 2.
\]

Here and in the sequel, \(P_{\rho}\) denotes probability computation when the true parameter is \(\rho\). The condition under which consistent discrimination rules do not exist turns out to be remarkably simple. Let \(f_\rho\) be the density of \(F_\rho\). We will show in Theorem 1 that \(\rho_1\) and \(\rho_2\) cannot be consistently discriminated if for \(\rho = \rho_1\) and \(\rho_2\),

\[
\int_0^1 \int_0^1 f^2_\rho(x, y)dx\,dy < \infty
\]

and the second largest singular value of the linear operator \(h \mapsto \int_0^1 f_\rho(x, \cdot)h(x)dx\), \(h \in L^2[0, 1]\), is strictly less than 1.

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

\[
F_n(x, y) = n^{-1}\left(\sum_{i=1}^{n} I_{[x_i \leq x]}, \sum_{i=1}^{n} I_{[y_i \leq y]}\right), \quad x, y \in [0, 1],
\]

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

\[
\text{cov}(Z_1(x_1), Z_1(x_2)) = x_1 \wedge x_2 - x_1x_2
\]

\[
\text{cov}(Z_2(y_1), Z_2(y_2)) = y_1 \wedge y_2 - y_1y_2
\]

\[
\text{cov}(Z_1(x), Z_2(y)) = F_\rho(x, y) - xy.
\]

Let \(P_{\rho}\), henceforth denote the probability distribution of the limiting Gaussian process \(Z\) described above under parameter value \(\rho\). Note that the standardization does not involve \(\rho\), so it is reasonable to argue that most of the information about \(\rho\) in \(F_n\) carries over to \(Z\). We also remark in passing that the weak convergence implies that \(\rho\) is identifiable in the broken sample setting so long as it is identifiable in the bivariate distribution \(F_\rho\). Suppose that for two given parameter values \(\rho_1\) and \(\rho_2\), \(P_{\rho_1}\) and \(P_{\rho_2}\) are equivalent, also called mutually absolutely continuous and denoted by \(P_{\rho_1} \equiv P_{\rho_2}\) here. Then it is clearly not possible to discriminate between
the two models with probability one based on $Z$. Theorem 3 shows that the same conditions of Theorem 1 plus some additional minor regularity condition ensure that $P_{ho_i} = P_{ho_n}, i = 1, 2$ and hence $P_{ho_1} = P_{ho_2}$.

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 $\rho$ can be consistently discriminated or even estimated.

2. Main results and examples

We assume that $F_\rho$ has a density $f_\rho$, and write

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

Define the linear operator

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

Suppose $A(\rho) < \infty$. Then $T_\rho$ 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 $T_\rho$ 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

$$f_\rho(x, y) = 1 + \sum_{i=1}^{\infty} \lambda_i,\rho \psi_i,\rho(x)\phi_i,\rho(y). \quad (2)$$
Thus,

\[ A(\rho) = 1 + \sum_{i=1}^{\infty} \lambda_{i,\rho}^2 < \infty. \]  

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 \( \rho_1 \) versus \( \rho_2 \) based on the broken sample.

**Remark 1.** The condition (HS) is 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 \( g_{n,\rho}(x, y) \) the density of the broken sample, i.e.

\[ g_{n,\rho}(x, y) = \frac{1}{n!} \sum_{\pi} f_{\rho}(x_{\pi}, y_{\pi}), \]

where the summation is taken over all permutations \( \pi \) of 1, \ldots, \( n \). By assumption, \( g_{n,\rho_0}(x, y) \equiv 1 \). As a result, \( g_{n,\rho}(x, y) \) can also be viewed as a likelihood ratio.

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

\[ \lim_{n \to \infty} P_{\rho_0}(g_{n,\rho}(X, Y) \leq x) = P(\xi \leq x) \quad \text{for all } x, \]

where

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

with the \( U_k, V_k \) denoting iid. standard normal random variables.

Observe that \( \log \xi \) is a constant plus a weighted average of independent \( \chi^2 \) random variables.

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

\[ f_{\rho,\delta} = \delta \wedge f_{\rho}, \quad \delta > 0, \]
Theorem 3. Suppose that the condition (HS) holds for some \( \rho \), and that each \( \delta > 0 \), \( f_{\rho,\delta} \) is square integrable in the Riemann sense on \([0, 1] \times [0, 1]\). Then \( \mathcal{P}_\rho = \mathcal{P}_{\rho_0} \) (see section 1 for notation). Thus, the class of probability distributions \( \mathcal{P}_\rho \) for which \( F_\rho \) 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 \( \mathcal{P}_{\rho_0} \) in terms of \( f_\rho \), where we show under the conditions stated in Theorem 3 that the Kulback-Leibler information between \( \mathcal{P}_\rho \) and \( \mathcal{P}_{\rho_0} \) 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 \( \rho \) and denote by \( \phi_\rho \) 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 \( \phi \) 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 \( f_\rho \) 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 \( \Phi \) and \( Q \) are the standard normal cdf and quantile function, respectively. It is easy to check that (HS) holds for each \( \rho \) 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 \( J_1 \) be iid. with \( P(J_1 = 1) = 1 - P(J_1 = 0) = \rho \) where \( \rho \in [0, 1] \) and

\[
X_i = J_i U_i + (1 - J_i) V_i, \quad Y_i = J_i U_i + (1 - J_i) W_i
\]

where \( U_i, V_i, W_i, 1 \leq i \leq n \) are iid; in this case, \( P_\rho(Y = X) = \rho \).
Example C. Let

\[ X_i = \rho U_i + (1 - \rho) V_i, \quad Y_i = \rho U_i + (1 - \rho) W_i \]  

(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 \( \rho \) 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 \( \rho \) 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 \( \rho \). Indeed the following can be proved.

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

\[ \hat{\rho}_n := \frac{1}{k_n} \sum_{i=1}^{k_n} \sum_{j=1}^{n} I(Y_j / X_i) \in (1 - \epsilon_n, 1 + \epsilon_n)) \to \rho \]

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

\[ f_\rho(x, y) = \rho^{-1} I(0 < x, y \leq \rho) + (1 - \rho)^{-1} I(\rho < x, y < 1), \quad x, y \in (0, 1). \]

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

\[ g(x) = \sqrt{\frac{1 - \rho}{\rho}} I(0 < x \leq \rho) - \sqrt{\frac{\rho}{1 - \rho}} I(\rho < x < 1). \]

Then \( \int_0^1 g(x)dx = 0, \int_0^1 g^2(x)dx = 1 \) and \( \int_0^1 \int_0^1 g(x) f_\rho(x, y) g(y) dy dx = 1 \) so that \( \lambda_{1,\rho} = 1 \). Consistent discrimination between any two distinct values \( \rho_1, \rho_2 \) is trivial; an obvious such rule is \( T_n = \rho_1 \) if \( \sum_{i=1}^{n} I(X_i \leq \rho_1) = \sum_{i=1}^{n} I(Y_i \leq \rho_1) \) 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 \( \rho \) 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 $\rho$. Then

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

**Proof.** Clearly,

$$E_{\rho,n} g_{n,\rho}^2 (X, Y) = \frac{1}{n!} \sum_{\pi} \int \int \prod_{i=1}^{n} f_{\rho}(x_i, y_{\pi_i}) dx_i dy_i = 1. \quad (5)$$

Also,

$$E_{\rho,n} g_{n,\rho}^2 (X, Y) = \frac{1}{(n!)^2} \sum_{\pi, \pi'} \int \int \prod_{i=1}^{n} f_{\rho}(x_i, y_{\pi_i}) f_{\rho}(x_i, y_{\pi'_i}) dx_i dy_i$$

$$= \frac{1}{(n!)^2} \sum_{\pi, \pi'} \int \int \prod_{i=1}^{n} \sum_{k=0}^{\infty} \lambda_k^2 \phi_k(y_{\pi_i}) \phi_k(y_{\pi'_i}) dy_i$$

$$= \frac{1}{n!} \sum_{\pi} \int \int \prod_{i=1}^{n} \sum_{k=0}^{\infty} \lambda_k^2 \phi_k(y_{\pi_i}) \phi_k(y_{\pi_i}) dy_i.$$

It is easy to verify that for a given permutation $\pi$

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

if the permutation $\pi$ consists of $\ell$ cycles of sizes $i_1, \ldots, i_\ell$ ($\sum i_i = n$). Therefore, if $j_i$ denotes the number of $t$ among $[i_1, \ldots, i_\ell]$, then

$$E_{\rho,n} g_{n,\rho}^2 (X, Y) = \sum_{\ell=1}^{n} \sum_{j_1+j_2++j_\ell=n} \prod_{i=1}^{\infty} \lambda_k^{2j_i}. \quad (6)$$

In fact, it is easy to see from this that $E_{\rho,n} g_{n,\rho}^2 (X, Y)$ 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
\[ E_{\rho_n G_{n, r}}(X, Y) = \prod_{k=1}^{\infty} (1 - \lambda_k^2)^{-1}. \]

Hence, \( E_{\rho_n G_{n, r}}(X, Y) \to \prod_{k=1}^{\infty} (1 - \lambda_k^2)^{-1}. \)

**Proof of Theorem 2.** Suppose that \( X_i, Y_i, i = 1, \ldots, n \) 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
\[
\frac{1}{n!} \sum_{\pi} f_\rho(X_i, Y_{\pi_i}) = \frac{1}{n!} \sum_{\pi} \prod_{i=1}^{n} (1 + W(X_i, Y_{\pi_i})) = 1 + \sum_{i=1}^{n} Q_i,
\]

where
\[
Q_i = \frac{(n-t)!}{n!} \sum_{1 \leq i_1 < \cdots < i_t \leq n} \sum_{1 \leq j_1 < \cdots < j_t \leq t} \prod_{i=1}^{t} W(X_{i_j}, Y_{\pi_{i_j}}).
\]

We first simplify \( Q_i \). Suppose \((s_1, \ldots, s_t)\) is a given ordered subset of \([1, \ldots, n]\). Since there are \((n-t)!\) different permutations such that \((\pi_{i_1}, \ldots, \pi_{i_t}) = (s_1, \ldots, s_t)\),
\[
Q_i = \frac{(n-t)!}{n!} \sum_{i_1 < \cdots < i_t \leq n} \sum_{1 \leq j_1 < \cdots < j_t \leq t} \prod_{i=1}^{t} W(X_{i_j}, Y_{\pi_{i_j}}).
\]

It is easy to verify that all \( Q_i \) are of mean 0 and uncorrelated with each other, and
\[
E Q_i^2 = \frac{1}{n!} \sum_{\pi} \prod_{j=1}^{t} \sum_{k=1}^{\infty} \lambda_k^2 \phi_k(Y_j) \phi_k(Y_{\pi_j}).
\]

Observe that the expectation in the above summation is 0 if \([s_1, \ldots, s_t] \neq [1, \ldots, t]\) due to the orthogonality of the functions \(\phi_k\)’s. Thus, for each non zero term, \([s_1, \ldots, s_t]\) can be considered as a permutation of \([1, \ldots, t]\). Classify the permutations by the numbers of cycles of sizes \(j = 1, \ldots, t\). Suppose the number of cycles of size \(j\) is \(v_j\), then we have \(v_1 + 2v_2 + \cdots + tv_t = t\) and
\[
E Q_i^2 = \frac{1}{t!} \sum_{v_1 + 2v_2 + \cdots + tv_t = t} \prod_{j=1}^{t} \frac{1}{v_j!} \left( \frac{1}{t} \sum_{k=1}^{\infty} \lambda_k^2 \right)^{v_j}. \]

Similar to the derivation of the limit of the left hand side of (6), we have
\[
E Q_i^2 = \frac{1}{2\pi it!} \int_{|z| = \rho} \frac{1}{z^{t+1}} \prod_{k=1}^{\infty} (1 - z \lambda_k^2)^{-1} dz.
\]

By choosing \( r \in (1, \lambda_1^{-2}) \), one sees that
\[
E Q_i^2 \leq \frac{1}{t!} r^{-t} \prod_{k=1}^{\infty} (1 - r \lambda_k^2)^{-1}.
\]
Thus, it follows that

$$\sum_{t=T}^{n} Q_t \to 0 \text{ in } L_2$$  \hfill (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 $\psi_k$ and $\phi_k$, it follows that for any fixed positive integer $m$, we have

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

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

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

and for $s \geq 3$,

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

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

$$Q_{mt} = \frac{(n-t)!}{n!} \sum_{1 \leq i_1 < \cdots < i_t \leq n} \sum_{j=1}^{t} W_m(X_{i_j}, Y_{i_j})$$

$$= \sum_{k_1, \ldots, k_r = 1}^{m} \prod_{j=1}^{t} \lambda_{k_j} \frac{(n-t)!}{n!} \sum_{1 \leq i_1 < \cdots < i_t \leq n} \sum_{j=1}^{t} \prod_{k_{i_j}}^{t} \psi_{k_j}(X_{i_j}) \phi_{k_j}(Y_{i_j})$$

$$= \sum_{k_1, \ldots, k_r = 1}^{m} \prod_{j=1}^{t} \lambda_{k_j} \frac{(n-t)!}{n!} \left( \sum_{i_1 < \cdots < i_t} \prod_{j=1}^{t} \psi_{k_j}(X_{i_j}) \right) \left( \sum_{i_1 < \cdots < i_t} \prod_{j=1}^{t} \phi_{k_j}(Y_{i_j}) \right).$$  \hfill (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

$$Q_{m1} = \sum_{k=1}^{m} \lambda_k n^{-1} \sum_{i=1}^{n} \psi_k(X_i) \sum_{i=1}^{n} \phi_k(Y_i) \overset{d}{\to} \sum_{k=1}^{m} \lambda_k U_k V_k$$
and by (8) and (9)

\[
Q_{m2} = \sum_{k_1, k_2 = 1}^{m} \lambda_{k_1} \lambda_{k_2} \frac{1}{2n(n-1)} \left[ \left( \sum_{i=1}^{n} \psi_{k_1}(X_i) \right) \left( \sum_{i=1}^{n} \psi_{k_2}(X_i) \right) - \sum_{i=1}^{n} \psi_{k_1}(X_i) \psi_{k_2}(X_i) \right] \\
\left[ \left( \sum_{i=1}^{n} \phi_{k_1}(Y_i) \right) \left( \sum_{i=1}^{n} \phi_{k_2}(Y_i) \right) - \sum_{i=1}^{n} \phi_{k_1}(Y_i) \phi_{k_2}(Y_i) \right] \\
\rightarrow \frac{1}{2} \sum_{k_1, k_2 = 1}^{m} \lambda_{k_1} \lambda_{k_2} (U_{k_1} U_{k_2} - \delta_{k_1, k_2}) (V_{k_1} V_{k_2} - \delta_{k_1, k_2}) \\
= \frac{1}{2} \left( \sum_{k=1}^{m} \lambda_{k} U_{k} V_{k} \right)^2 - \sum_{k=1}^{m} \lambda_{k}^2 (U_{k}^2 + V_{k}^2 - 1) .
\]

For \( t = 3 \), we have

\[
n^{-3/2} \sum_{i_1, i_2, i_3} \psi_{i_1}(X_{i_1}) \psi_{i_2}(X_{i_2}) \psi_{i_3}(X_{i_3}) \\
= n^{-3/2} \prod_{i=1}^{n} \sum_{j=1}^{n} \psi_{i}(X_{i}) - n^{-3/2} \sum_{i=1}^{n} \psi_{i}(X_{i}) \psi_{i}(X_{i}) \sum_{i=1}^{n} \psi_{i}(X_{i}) \psi_{i}(X_{i}) \\
- n^{-3/2} \sum_{i=1}^{n} \psi_{i}(X_{i}) \psi_{i}(X_{i}) \sum_{i=1}^{n} \psi_{i}(X_{i}) \psi_{i}(X_{i}) - n^{-3/2} \sum_{i=1}^{n} \psi_{i}(X_{i}) \psi_{i}(X_{i}) \sum_{i=1}^{n} \psi_{i}(X_{i}) \\
+ n^{-3/2} \sum_{i=1}^{n} \psi_{i}(X_{i}) \psi_{i}(X_{i}) \psi_{i}(X_{i}) \\
\rightarrow U_{k_1} U_{k_2} U_{k_3} - \delta_{k_1, k_2} U_{k_1} - \delta_{k_2, k_3} U_{k_2} - \delta_{k_3, k_1} U_{k_3}.
\]

Similarly,

\[
n^{-3/2} \sum_{x_1, x_2, x_3} \phi_{x_1}(Y_{x_1}) \phi_{x_2}(Y_{x_2}) \phi_{x_3}(Y_{x_3}) \\
\rightarrow V_{k_1} V_{k_2} V_{k_3} - \delta_{k_1, k_2} V_{k_1} - \delta_{k_2, k_3} V_{k_2} - \delta_{k_3, k_1} V_{k_1}.
\]

Therefore,

\[
Q_{m3} \rightarrow \frac{1}{6} \left[ \left( \sum_{k=1}^{m} \lambda_{k} U_{k} V_{k} \right)^3 - 3 \left( \sum_{k=1}^{m} \lambda_{k} U_{k} V_{k} \right) \left( \sum_{k=1}^{m} \lambda_{k}^2 (U_{k}^2 + V_{k}^2 - 1) \right) \\
+ 6 \sum_{k=1}^{m} \lambda_{k}^3 U_{k} V_{k} \right] .
\]
In general, by the inclusion-exclusion identity and noticing (10), we have for a given $k = (k_1, \cdots, k_t)$,

$$n^{-t/2} \sum_{i_1, \cdots, i_t} \prod_{j=1}^t \psi_{k_j}(X_{i_j})$$

$$d \sum_{j=1}^t \prod_{j \neq j_1, j_2} \prod_{j \notin \{j_1, j_2\}} U_{k_j}$$

$$+ \sum_{\{j_1, j_2\} \subset \{1, \cdots, t\}} \prod_{j \notin \{j_1, j_2\}} U_{k_j}$$

$$- \sum_{\{j_1, j_2\} \subset \{1, \cdots, t\}} \prod_{j \notin \{j_1, j_2\}} U_{k_j} + \cdots$$

$$=: S_{t,k,1} - S_{t,k,2} + \cdots + (-1)^{t/2} S_{t,k,[t/2]+1},$$

where the sum in $S_{t,k,1+\ell}, \ell \geq 1$, runs over all possible $\ell$ pairs of indices $\{\{j_1, j_2\}, \cdots, \{j_{2\ell-1}, j_{2\ell}\}\}$, in which the $2\ell$ 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

$$n^{-t/2} \sum_{s_1, \cdots, s_t} \prod_{j=1}^t \psi_{k_j}(Y_{s_j})$$

$$d \sum_{j=1}^t \prod_{j \neq j_1, j_2} \prod_{j \notin \{j_1, j_2\}} V_{k_j}$$

$$+ \sum_{\{j_1, j_2\} \subset \{1, \cdots, t\}} \prod_{j \notin \{j_1, j_2\}} V_{k_j}$$

$$- \sum_{\{j_1, j_2\} \subset \{1, \cdots, t\}} \prod_{j \notin \{j_1, j_2\}} V_{k_j} + \cdots$$

$$=: T_{t,k,1} - T_{t,k,2} + \cdots + (-1)^{t/2} T_{t,k,[t/2]+1},$$

Substituting these two limits into (11), we get

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

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

$$
\sum_{k_1, \ldots, k_t}^m \prod_{j=1}^t \lambda_{k_j} S_{t,k_1} T_{t,k_1}
$$

$$
= \sum_{k_1, \ldots, k_t}^m \prod_{j=1}^t \lambda_{k_j} \prod_{j=1}^t U_{k_j} V_{k_j}
$$

$$
= \left( \sum_{k_1}^m \lambda_k U_k V_k \right)^t \to \left( \sum_{k_1}^\infty \lambda_k U_k V_k \right)^t \text{ as } m \to \infty. \tag{13}
$$

Next,

$$
\sum_{k_1, \ldots, k_{t+1}}^m \prod_{j=1}^t \lambda_{k_j} S_{t,k_1} T_{t,k_1}
$$

$$
= \sum_{\cup_{j=1}^t \{j_{2s-1}, j_{2s}\} \subset \{1, \ldots, t\}} \sum_{k_1, \ldots, k_t}^m \prod_{j=1}^t \lambda_{k_j} U_{k_j} \prod_{j=1}^t \delta_{k_{2s-1}, k_{2s}} \prod_{j \not\in \{1, \ldots, t_1\}} V_{k_j}
$$

$$
= \frac{t!}{2^t! (t-2t)!} \left( \sum_{k_1}^m \lambda_k^2 U_k \right)^t \left( \sum_{k_1}^m \lambda_k^2 U_k V_k \right)^{t-2t}
$$

$$
\to \frac{t!}{2^t! (t-2t)!} \left( \sum_{k_1}^\infty \lambda_k^2 U_k \right)^t \left( \sum_{k_1}^\infty \lambda_k^2 U_k V_k \right)^{t-2t}. \tag{14}
$$

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

$$
\sum_{k_1, \ldots, k_{t+1}}^m \prod_{j=1}^t \lambda_{k_j} S_{t,k_1} T_{t,k_1}
$$

$$
= \sum_{\cup_{j=1}^t \{j_{2s-1}, j_{2s}\} \subset \{1, \ldots, t\}} \sum_{\cup_{j=1}^t \{j_{2s-1}, j_{2s}\} \subset \{1, \ldots, t\}} \sum_{k_1, \ldots, k_t}^m \prod_{j=1}^t \lambda_{k_j} \prod_{j=1}^t \delta_{k_{2s-1}, k_{2s}} \prod_{j \not\in \{1, \ldots, t_1\}} V_{k_j}
$$

$$
\prod_{j \not\in \{1, \ldots, t_1\}} U_{k_j} \prod_{\phi(1, \ldots, t_2)} V_{k_j}. \tag{15}
$$

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 $\{1, 2, \ldots, t\}$. It is said to be a

- a cycle of length $h$ (even) if $\{v_1, v_2, \ldots, v_{h-1}, v_h\} \subset \{f_1, f_2, \ldots, f_{2h-1}, f_{2h}\}$ and $\{v_{h-2}, v_{h-1}, v_h, v_1\} \subset \{f_1, f_2, \ldots, f_{2h-1}, f_{2h}\}$, where $(v_1, v_2, \ldots, v_h)$ is any permutation of $(u_1, u_2, \ldots, u_h)$.

- a $uu$-chain of length $h$ (even) if $\{u_1, u_2, \ldots, u_{h-1}, u_h\} \subset \{f_1, f_2, \ldots, f_{2h-1}, f_{2h}\}$ and $\{u_{h-2}, u_{h-1}, u_h, u_1\} \subset \{f_1, f_2, \ldots, f_{2h-1}, f_{2h}\}$ and $u_1, u_h \not\in \{i_1, \ldots, i_{2h}\}$.
· a uu-chain of length h (even) if \{u_2, u_3, \ldots, u_{h-2}, u_{h-1}\} \subset \{j_1, j_2, \ldots, j_{2\ell-1}, j_{2\ell}\} and \{u_1, u_2, \ldots, u_{h-1}, u_h\} \subset \{i_1, i_2, \ldots, i_{2\ell-1}, i_{2\ell}\} and 

u_1, u_h \notin \{j_1, \ldots, j_{2\ell}\}.

· a vv-chain of length h (odd) if \{u_1, u_2, \ldots, u_{h-2}, u_{h-1}\} \subset \{j_1, j_2, \ldots, j_{2\ell-1}, j_{2\ell}\} and \{u_2, u_3, \ldots, u_{h-1}, u_h\} \subset \{i_1, i_2, \ldots, i_{2\ell-1}, i_{2\ell}\} and 

u_1, u_h \notin \{j_1, \ldots, j_{2\ell}\}, or alternatively.

An integer u \leq t is called a singleton if u \notin \{j_1, \ldots, j_{2\ell}\} \cup \{i_1, \ldots, 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 uu-chain of length 1. In the sequel, we shall not specify singletons.

Observe that for each given partitions \{\{j_1, j_2\}, \ldots, \{j_{2\ell-1}, j_{2\ell}\}\} and \{\{i_1, i_2\}, \ldots, \{i_{2\ell-1}, i_{2\ell}\}\}, and we can uniquely partitioned into disjoint sets each of which is a cycle or contains a uu, vv or a uu-uu chain. As a simple illustration, let \ell = 2 and consider the partitions

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

and

\{\{i_1, i_2\}, \ldots, \{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 uu-chain (which is also a singleton), and (10, 12, 11) is a uu-uu chain. Then it is easy to see that for this example,

\[
\sum_{k_1, \ldots, k_\ell} \prod_{x=1}^{\ell} \lambda_{k_x} \prod_{s'=1}^{\ell'} \delta_{k_{2s'-1}, k_{2s'}} \prod_{s'=1}^{\ell'} \delta_{k_{2s'-1}, k_{2s'}} \prod_{j \notin \{j_1, \ldots, j_{2\ell}\}} U_{k_j} \prod_{i \notin \{i_1, \ldots, i_{2\ell}\}} V_{k_i} 
= \left( \sum_{k=1}^{m} \lambda_k^2 \right) \left( \sum_{k=1}^{m} \lambda_k^4 V_k \right) \left( \sum_{k=1}^{m} \lambda_k^2 U_k \right) \left( \sum_{k=1}^{m} \lambda_k^4 V_k \right).
\]

Observe that each cycle, vv, uu and uu-uu chain of length h produce factors \sum_{k=1}^{m} \lambda_k^h, \sum_{k=1}^{m} \lambda_k^h V_k, \sum_{k=1}^{m} \lambda_k^h U_k 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\}, \ldots, \{j_{2\ell-1}, j_{2\ell}\}\} and \{\{i_1, i_2\}, \ldots, \{i_{2\ell-1}, i_{2\ell}\}\}, denote by \nu_{h1}, the number of cycles of length h, and \nu_{h2}, \nu_{h3} and \nu_{h4} the number of uu, vv and uu-uu chains of length h, respectively. Note that

\[
\sum_{h=1}^{t} h \nu_{h1} = t.
\]

Then

\[
\sum_{k_1, \ldots, k_\ell} \prod_{x=1}^{\ell} \lambda_{k_x} \prod_{s'=1}^{\ell'} \delta_{k_{2s'-1}, k_{2s'}} \prod_{s'=1}^{\ell'} \delta_{k_{2s'-1}, k_{2s'}} \prod_{j \notin \{j_1, \ldots, j_{2\ell}\}} U_{k_j} \prod_{i \notin \{i_1, \ldots, i_{2\ell}\}} V_{k_i} 
= \prod_{h \text{ even}} \left( \sum_{k=1}^{m} \lambda_k^h \right)^{\nu_{h1}} \left( \sum_{k=1}^{m} \lambda_k^h V_k \right)^{\nu_{h2}} \left( \sum_{k=1}^{m} \lambda_k^h U_k \right)^{\nu_{h3}} \prod_{h \text{ odd}} \left( \sum_{k=1}^{m} \lambda_k^h U_k \right)^{\nu_{h4}}.
\]

In addition to \sum_{h=1}^{t} h \nu_{h1} = t, we should also have the constraints \nu_{h1} = \nu_{h2} = \nu_{h3} = 0 for odd h, \nu_{h4} = 0 for even h, and \nu_{h1}(2\nu_{h3} + \nu_{h4}) = t - 2\ell. These constraints come directly from the definition; for example, each uu-chain and vv-chain correspond to one and two elements in 1, \ldots, t, respectively, that are not in \{\{i_1, i_2\}, \ldots, \{i_{2\ell-1}, i_{2\ell}\}\}.
While each \(j-i\) combination, namely, \(\{j_1, j_2, \ldots, j_s\} \text{ and } \{i_1, i_2, \ldots, i_s\}\), 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 \(v_{hi}, h = 1, \ldots, t, i = 1, 2, 3, 4\) partition \(\{1, 2, \ldots, t\}\) into groups \(G_{hi}^{(s)}\) of sizes \(h, s = 1, 2, \ldots, v_{hi}, \ h = 1, \ldots, t, i = 1, 2, 3, 4\) The number of such partitions is

\[
\frac{t!}{\prod_{h=1}^{t-4} v_{hi}! (h!)^{v_{hi}}}
\]

For a given partition of \(G_{hi}^{(s)}\), we compute the number of possible \(j-i\) pairs, namely, sets \(\{j_1, j_2, \ldots, j_s\} \text{ and } \{i_1, i_2, \ldots, i_s\}\), using elements of \(G_{hi}^{(s)}\).

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

Next consider any \(G_{hi}^{(s)}\) with \(h \geq 2\) even. For each permutation \((u_1, \ldots, u_h)\) of elements of \(G_{hi}^{(s)}\), we construct the \(i-j\) pairs by putting \([u_1, u_2], \ldots, [u_{h-1}, u_h]\) into the \(i\)-pairs, \([u_2, u_3], \ldots, [u_{h-2}, u_{h-1}]\) 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 \(G_{hi}^{(s)}\) with \(h \geq 2\) even, we have \(h!/2\) ways to construct all possible \(j-i\) pairs.

Finally consider any \(G_{hi}^{(s)}\) with \(h \geq 2\) odd. For each permutation \((u_1, \ldots, u_h)\) of elements of \(G_{hi}^{(s)}\), we construct the \(i-j\) pairs by putting \([u_1, u_2], \ldots, [u_{h-2}, u_{h-1}]\) into the \(i\)-pairs and \([u_2, u_3], \ldots, [u_{h-1}, u_h]\) 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 \(G_{hi}^{(s)}\) as both unpaired \(j\)’s as well as the unpaired \(i\)’s. Combining these, by (15), we have

\[
\sum_{k_1, \ldots, k_t = 1}^{m} \prod_{j=1}^{t} \lambda_{k_j} \delta_{s, k_j} \delta_{t, k_j + 1} T_{i, k_j, k_j + 1}
= \sum_{h \text{ even}} \lambda_{k} \delta_{h} \delta_{h+1} \prod_{h \text{ even}} \frac{1}{v_{h1}! v_{h2}! v_{h3}!} \left( \frac{1}{h} \sum_{k=1}^{m} \lambda_{k} \right)_{v_{h1}} \left( \frac{1}{2} \sum_{k=1}^{m} \lambda_{k} V_{k} \right)_{v_{h2}}
\]

\[
\left( \frac{1}{2} \sum_{k=1}^{m} \lambda_{k} U_{k} \right)_{v_{h3}} \prod_{h \text{ odd}} \frac{1}{v_{h4}!} \left( \sum_{k=1}^{\infty} \lambda_{k} U_{k} V_{k} \right)_{v_{h4}}
\]

\[
\to \sum_{h \text{ even}} \lambda_{k} \delta_{h} \delta_{h+1} \prod_{h \text{ even}} \frac{1}{v_{h1}! v_{h2}! v_{h3}!} \left( \frac{1}{h} \sum_{k=1}^{m} \lambda_{k} \right)_{v_{h1}} \left( \frac{1}{2} \sum_{k=1}^{m} \lambda_{k} V_{k} \right)_{v_{h2}}
\]

\[
\left( \frac{1}{2} \sum_{k=1}^{m} \lambda_{k} U_{k} \right)_{v_{h3}} \prod_{h \text{ odd}} \frac{1}{v_{h4}!} \left( \sum_{k=1}^{\infty} \lambda_{k} U_{k} V_{k} \right)_{v_{h4}}
\]
\[
\left( \frac{1}{2} \sum_{k=1}^{\infty} \lambda_k^h U_k^2 \right)^{v_{h3}} \prod_{h \text{ odd}} \frac{1}{v_{h4}} \left( \sum_{k=1}^{\infty} \lambda_k^h U_k V_k \right)^{v_{h4}},
\]

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

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
\[
(-1)^{\ell + \ell'} = (-1)^{t + \sum_{h} (v_{h2} + v_{h3} + v_{h4})}.
\]

Thus,
\[
Q_t \overset{d}{\to} (-1)^{t} \sum^* \prod_{h \text{ even}} \frac{1}{v_{h1}!v_{h2}!v_{h3}!} \left( \frac{1}{2} \sum_{k=1}^{\infty} \lambda_k^h U_k^2 \right)^{v_{h1}} \left( -\frac{1}{2} \sum_{k=1}^{\infty} \lambda_k^h V_k^2 \right)^{v_{h2}} \prod_{h \text{ odd}} \frac{1}{v_{h4}} \left( -\sum_{k=1}^{\infty} \lambda_k^h U_k V_k \right)^{v_{h4}},
\]

where \( \sum^* \) runs over all possible integers \( v_{hi} \) subject to \( \sum_{h,t} h v_{hi} = t \), \( v_{h1} = v_{h2} = v_{h3} = 0 \) for odd \( h \), \( v_{h4} = 0 \) for even \( h \). We further simplify it as
\[
Q_t \overset{d}{\to} (-1)^{t} \sum^{***} \prod_{h \text{ even}} \frac{1}{v_{h0}!} \left( \sum_{k=1}^{\infty} \lambda_k^h \left( \frac{1}{h} U_k^2 - \frac{1}{2} V_k^2 \right) \right)^{v_{h0}} \prod_{h \text{ odd}} \frac{1}{v_{h4}} \left( -\sum_{k=1}^{\infty} \lambda_k^h U_k V_k \right)^{v_{h4}} =: \xi_t,
\]

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

Note that \( \xi_t \) is the coefficient of \( z^t \) in the Taylor expansion of
\[
H(z) = \exp \left( \sum_{h=0}^{\infty} \frac{(-1)^h}{h!} \sum_{k=1}^{\infty} \lambda_k^h \left( \frac{1}{h} U_k^2 - \frac{1}{2} V_k^2 \right) \right).
\]

Straightforward algebra shows that
\[
\sum \xi_t = H(1) = \prod_{k=1}^{\infty} (1 - \lambda_{k,R}^2)^{-1/2} \exp \left( -\frac{1}{2} \sum_{k=1}^{\infty} \frac{\lambda_{k,R}^2}{1 - \lambda_{k,R}^2} (U_k^2 + V_k^2) + \sum_{k=1}^{\infty} \frac{\lambda_{k,R}^2}{1 - \lambda_{k,R}^2} U_k V_k \right).
\]

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 \( \xi \) described in the theorem. This concludes the proof. \( \square \)
3.2. Proof of Theorem 1

Let \( T_n \) be a consistent discrimination rule for \( \rho_1 \) versus \( \rho_2 \). Write, for any fixed \( M \in (0, \infty) \),

\[
P_{\rho_1}(T_n = \rho_1) = P_{\rho_1}(T_n = \rho_1, g_{n, \rho_1}(X, Y) > M g_{n, \rho_2}(X, Y)) + P_{\rho_1}(T_n = \rho_1, g_{n, \rho_1}(X, Y) \leq M g_{n, \rho_2}(X, Y)).
\]

Suppose that \( P_{\rho_1}(g_{n, \rho_1}(X, Y) > M g_{n, \rho_2}(X, Y)) \neq 1 \). Since \( P_{\rho_1}(T_n = \rho_1) \to 1 \), we have \( P_{\rho_1}(T_n = \rho_1, g_{n, \rho_1}(X, Y) \leq M g_{n, \rho_2}(X, Y)) \to 0 \). Thus

\[
\limsup_{n \to \infty} P_{\rho_2}(T_n = \rho_1) \geq \limsup_{n \to \infty} M^{-1} P_{\rho_1}(T_n = \rho_1, g_{n, \rho_1}(X, Y) \leq M g_{n, \rho_2}(X, Y)) > 0.
\]

This contradicts the assumption that \( T_n \) is consistent. Hence, we must have

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

Now,

\[
E_{\rho_1, \nu_{n, \rho_1}}(X, Y) \geq \int g_{n, \rho_1}^2(x, y) I(g_{n, \rho_1}(x, y) > M g_{n, \rho_2}(x, y)) dxdy \tag{17}
\]

\[
\geq M \int g_{n, \rho_1}(x, y) g_{n, \rho_2}(x, y) I(g_{n, \rho_1}(x, y) > M g_{n, \rho_2}(x, y)) dxdy
\]

\[
\geq M d \int g_{n, \rho_1}(x, y) I(g_{n, \rho_1}(x, y) > M g_{n, \rho_2}(x, y), g_{n, \rho_2}(x, y) > d) dxdy
\]

\[
\geq M d \int g_{n, \rho_1}(x, y) I(g_{n, \rho_1}(x, y) > M g_{n, \rho_2}(x, y)) dxdy
\]

\[
-M d \int g_{n, \rho_1}(x, y) I(g_{n, \rho_2}(x, y) \leq d) dxdy.
\]

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 \( \xi \) has a continuous distribution,

\[
P_{\rho_0}(g_{n, \rho_2}(X, Y) \leq d) \to 0.
\]

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

\[
\int g_{n, \rho_1}(x, y) I(g_{n, \rho_2}(x, y) \leq d) dxdy \to 0. \tag{18}
\]

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

\[
E_{\rho_0} g_{n, \rho_1}^2(X, Y) \to \infty.
\]

This contradicts Lemma 5 and concludes the proof. \( \square \)
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 $\sigma$-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 $\sigma$-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

$$\lim_{m \to \infty} \min_{1 \leq i \leq m} \max_{1 \leq j \leq m} m \int_{x=\frac{i}{m}}^{\frac{i+1}{m}} \int_{y=\frac{j}{m}}^{\frac{j+1}{m}} f_\rho(x, y) dx dy = 0. \quad (19)$$

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

$$\min_{1 \leq i \leq m} \max_{1 \leq j \leq m} m \int_{x=\frac{i}{m}}^{\frac{i+1}{m}} \int_{y=\frac{j}{m}}^{\frac{j+1}{m}} f_\rho(x, y) dx dy \leq \frac{\delta_m}{m} + \frac{m}{\delta_m} \min_{i} \int_{(i-1)/m}^{i/m} \int_{0}^{1} f^2_\rho(x, y) dx dy. \quad (20)$$

By the condition (HS),

$$\min_{i} \int_{(i-1)/m}^{i/m} \int_{0}^{1} f^2_\rho(x, y) dx dy \to 0.$$

Thus, we may select a sequence $\delta_m$ 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 $i_m \in \{1, \ldots, m\}$ such that

$$\lim_{m \to \infty} \max_{1 \leq j \leq m} m P\left(\frac{i_m - 1}{m} < X \leq \frac{i_m}{m}, \frac{i - 1}{m} < Y \leq \frac{i}{m}\right) = 0. \quad (21)$$

We will apply Theorem 6 by letting $\mathcal{F}_m = \sigma\{Z_1(\frac{i-1}{m}), Z_1(\frac{i}{m}), Z_2(\frac{i-1}{m}) - Z_2(\frac{i}{m}) : i \in \{1, \ldots, m\} - \{i_m\}\}$ 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

$$U = \left(Z_1\left(\frac{i-1}{m}\right) - Z_1\left(\frac{i}{m}\right), Z_2\left(\frac{i-1}{m}\right) - Z_2\left(\frac{i}{m}\right) : i \in \{1, \ldots, m\} - \{i_m\}\right)$$

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

\[\Sigma_{m,\rho} = \begin{bmatrix} A_m & B_{m,\rho} \\ B_{m,\rho}^T & A_m \end{bmatrix}\]

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

\[
B_{m,\rho} = \left[ \begin{array}{c} \mathcal{P}_\rho \left( \frac{i-1}{m} < X \leq \frac{i}{m}, \frac{j-1}{m} < Y \leq \frac{j}{m} \right) \\ -\frac{1}{m^2}, i, j \in \{1, \ldots, m\} - \{i_m\} \end{array} \right],
\]

where \(\mathbf{I}\) 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 \(\mathcal{P}_{\rho_0}\) with respect to \(\mathcal{F}_m\) is

\[
J_m = \frac{1}{2}\{\text{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho}) + \text{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho_0}) - 4(m-1)\}.
\]

By assumption \(B_{m,\rho_0} = 0\) and hence

\[
\text{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho}) = 2(m-1).
\]

As a result,

\[
J_m = \frac{1}{2}\text{tr}(\Sigma_{m,\rho}^{-1}\Sigma_{m,\rho}) - (m-1)
\]

which, by Lemma 8 below, is equal to

\[
\text{tr}\{I_{m-1} - A_m^{-1}B_{m,\rho}A_m^{-1}B_{m,\rho}^T\} - (m-1) = \sum_{i=1}^{m-1}(1 - \lambda_{m,i}^2)^{-1} - (m-1)
\]

\[
= \sum_{i=1}^{m-1}\frac{\lambda_{m,i}^2}{1 - \lambda_{m,i}^2},
\]

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

\[
J_m = \sum_{i=1}^{\infty}\frac{\lambda_{i}^2}{1 - \lambda_{i}^2} < \infty
\]

where \(\lambda_{i} = \lambda_{i,\rho}\). By Theorem 6, we have \(\mathcal{P}_\rho \equiv \mathcal{P}_{\rho_0}\). This concludes the proof. \(\square\)
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

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

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

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

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

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

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

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

$$= mM \begin{bmatrix} \frac{m}{m-1}V_{B_{m,\rho}1} & \sqrt{\frac{m}{m-1}V_{B_{m,\rho}M_1}} \\ \sqrt{\frac{m}{m-1}M_1'B_{m,\rho}1} & M_1'B_{m,\rho}M_1 \end{bmatrix} M',$$

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

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

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

$$\sum_{i=1}^{m-1} (\lambda_{m,i} - n_{m,i})^2 \leq m^2 \left( [V_{B_{m,\rho}1}]^2 + \frac{(\sqrt{m} - 1)^2}{m-1}V_{B_{m,\rho}M_1}M_1'B_{m,\rho}1 \right)$$

$$\leq m^2 \left( [V_{B_{m,\rho}1}]^2 + V_{B_{m,\rho}B_{m,\rho}'}1 \right).$$
The broken sample problem

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

\[ I B_{m, \rho} = \text{cov} \left( I \left( \frac{i_m - 1}{m} < X \leq \frac{i_m}{m} \right), I \left( \frac{j - 1}{m} < Y \leq \frac{j}{m} \right) \right) \]

\[ = P \left( \frac{i_m - 1}{m} < X \leq \frac{i_m}{m}, \frac{i_m - 1}{m} < Y \leq \frac{i_m}{m} \right) - m^{-2} = o(m^{-1}). \]

Similarly,

\[ I B_{m, \rho} = \sum_{j \neq i_m} \text{cov} \left( I \left( \frac{i_m - 1}{m} < X \leq \frac{i_m}{m} \right), I \left( \frac{j - 1}{m} < Y \leq \frac{j}{m} \right) \right) \]

\[ \leq 2 \sum_{j \neq i_m} \left[ P^2 \left( \frac{i_m - 1}{m} < X \leq \frac{i_m}{m}, \frac{j - 1}{m} < Y \leq \frac{j}{m} \right) + m^{-4} \right] \]

\[ \leq 2o(m^{-1}) \sum_{j \neq i_m} P \left( \frac{i_m - 1}{m} < X \leq \frac{i_m}{m}, \frac{j - 1}{m} < Y \leq \frac{j}{m} \right) + o(m^{-3}) \]

\[ = o(m^{-2}) \]

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 \( i \geq m \). Let the singular decomposition of \( m B_{m, \rho} \) be given by

\[ m B_{m, \rho} = \sum_{1 \leq i \leq m} \eta_{m,i} \psi_{m,i} \phi_{m,i}, \quad (25) \]

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

\[ p_m(x, y) = \sum_{1 \leq i, j \leq m, i \neq j} (\xi_{i,j} - 1)I_{\left[ \frac{i-1}{m}, \frac{i}{m} \right]} \times I_{\left[ \frac{j-1}{m}, \frac{j}{m} \right]}(x, y), \quad (26) \]

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

\[ \psi_{m,i}(x) = \sqrt{m} \sum_{1 \leq k \leq m, k \neq i} \psi_{m,i,k} I_{\left[ \frac{i-1}{m}, \frac{i}{m} \right]}(x), \]

\[ \phi_{m,i}(y) = \sqrt{m} \sum_{1 \leq k \leq m, k \neq i} \phi_{m,i,k} I_{\left[ \frac{j-1}{m}, \frac{j}{m} \right]}(y). \]
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 \( T_{pm} : g \rightarrow \int p_m(\cdot, y)g(y)dy \) as
\[ p_m(x, y) = \sum_{1 \leq i \leq m \atop i \neq i_m} \eta_{m,i} \psi_{m,i}(x)\phi_{m,i}(y) \]

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
\[ \sum_{i=1}^{\infty} (\eta_{m,i} - \lambda_i)^2 \]
\[ \leq \int_0^1 \int_0^1 (p_m(x, y) - f_{\rho}(x, y) + 1)^2dxdy \]
\[ = \int_0^1 \int_0^1 I(x \text{ or } y \in \left[ \frac{i_m - 1}{m}, \frac{i_m}{m} \right]) (f_{\rho}(x, y) - 1)^2dxdy \]
\[ + \sum_{1 \leq i, j \leq m \atop i \neq j \neq i_m} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (\xi_{i,j} - f_{\rho}(x, y))^2dxdy. \]

The first term on the right-hand side tends to 0 as \( m \) tends to \( \infty \) since \( f_{\rho} \) 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
\[ B_{m,\delta,1} = \sum_{1 \leq i, j \leq m \atop i \neq j \neq i_m} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (\xi_{i,j,\delta} - f_{\rho,\delta}(x, y))^2dxdy, \]
\[ B_{m,\delta,2} = \sum_{1 \leq i, j \leq m \atop i \neq j \neq i_m} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (\xi_{i,j} - \xi_{i,j,\delta})^2dxdy \]
\[ B_{m,\delta,3} = \sum_{1 \leq i, j \leq m \atop i \neq j \neq i_m} \int_{\frac{i-1}{m}}^{\frac{i}{m}} \int_{\frac{j-1}{m}}^{\frac{j}{m}} (f_{\rho}(x, y) - f_{\rho,\delta}(x, y))^2dxdy. \]
The broken sample problem

For each fixed \( \delta \), \( B_{m,\delta,1} \to 0 \) as \( m \) tends to \( \infty \) by Riemann integrability of \( f_{x,\delta}^2 \).

Applying the Cauchy-Schwarz inequality,

\[
B_{m,\delta,2} = m^2 \sum_{1 \leq i, j \leq m} \left( \int_{i/m}^{(i+1)/m} \int_{j/m}^{(j+1)/m} (f_{x,y} - f_{x,\delta}(x,y)) \, dx \, dy \right)^2 \\
\leq \sum_{1 \leq i, j \leq m} \int_{i/m}^{(i+1)/m} \int_{j/m}^{(j+1)/m} (f_{x,y} - f_{x,\delta}(x,y))^2 \, dx \, dy = B_{m,\delta,3}.
\]

Now,

\[
B_{m,\delta,3} = \sum_{1 \leq i, j \leq m} \int_{i/m}^{(i+1)/m} \int_{j/m}^{(j+1)/m} (f_{x,y} - f_{x,\delta}(x,y))^2 \, dx \, dy \\
\leq \int_0^1 \int_0^1 (f_{x,y} - f_{x,\delta}(x,y))^2 \, dx \, dy.
\]

Since \( 0 \leq f_{x,y} - f_{x,\delta}(x,y) \leq f_{x,y} \), 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 concludes 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 \( \lambda_{m,i} \leq \lambda_i + \epsilon < 1 \) for all \( i \) and all large \( m \). Thus it is straightforward to conclude from (23) and (3) that that

\[
\sum_{i=1}^{m-1} \frac{\lambda_{m,i}^2}{1 - \lambda_{m,i}^2} \to \sum_{i=1}^{\infty} \frac{\lambda_i^2}{1 - \lambda_i^2} \leq (1 - \lambda_1)^{-1} (A(\rho) - 1) < \infty.
\]

This concludes the proof. \( \square \)

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

\[
\text{tr}(\Sigma_{m,\rho}^{-1} \Sigma_{m,\rho}) = 2\text{tr}([I_{m(m)} - A_m^{-1} B_{m,\rho} A_m^{-1} B_m']^{-1}).
\]

**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 \( \rho \) in \( A_m \) and \( B_{m,\rho} \). By Theorem 8.5.11 of Harvill (1997),

\[
\Sigma_{m,\rho}^{-1} = \begin{bmatrix} A & B \\ B' & A \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1} B (A - B' A^{-1} B)^{-1} B' A^{-1} - A^{-1} B (A - B' A^{-1} B)^{-1} \\ -(A - B' A^{-1} B)^{-1} B' A^{-1} (A - B' A^{-1} B)^{-1} \end{bmatrix}.
\]
Hence
\[
\Sigma_{m, \rho}^{-1} \Sigma_{m, \rho_0} = \begin{bmatrix}
A & B \\
B' & A
\end{bmatrix}^{-1} \begin{bmatrix}
A & 0 \\
0 & A
\end{bmatrix}
= \begin{bmatrix}
I_n + A^{-1}B(A - B'A^{-1}B)^{-1}B' - A^{-1}B(A - B'A^{-1}B)^{-1}A \\
-(A - B'A^{-1}B)^{-1}B' & (A - B'A^{-1}B)^{-1}A
\end{bmatrix}.
\]

Thus,
\[
\text{tr}(\Sigma_{m, \rho}^{-1} \Sigma_{m, \rho_0})
= \text{tr}(I_n + A^{-1}B(A - B'A^{-1}B)^{-1}B' + (A - B'A^{-1}B)^{-1}A)
= \text{tr}(I_n) + \text{tr}([A - B'A^{-1}B]^{-1}(B'A^{-1}B - A)] + 2\text{tr}([A - B'A^{-1}B]^{-1}A)]
= 2\text{tr}([A - B'A^{-1}B]^{-1}A]
= 2\text{tr}([I_n - A^{-1}B(A - B')^{-1}A)]
\]

\[\square\]

3.4. Proof of Theorem 4

First,
\[
A(\rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w_\rho^2(x, y)/(h(x)h(y))dx dy
\]
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,
\[
A(\rho) \geq c \sum_{j \neq 0} \int_{j}^{j+1} \int_{j}^{j+1} \frac{1}{j^4 h(x)h(y)} dx dy = \infty.
\]
Next we prove the consistency of the estimator \(\hat{\rho}_n\). For convenience, drop \(n\) in \(k_n\) and \(\varepsilon_n\). Let \(i^\ast\) be the index of \(X(i)\). Write
\[
\hat{\rho}_n := S_n + R_n
\]
where \(S_n = \frac{1}{k} \sum_{i=1}^{k} I(Y_{i^\ast}/X(i) \in (1 - \varepsilon, 1 + \varepsilon))\)
The broken sample problem

and

\[ R_n = \frac{1}{k} \sum_{i=1}^{k} \sum_{1 \leq j \leq n, j \neq i^*} I(Y_j / X(i) \in (1 - \varepsilon, 1 + \varepsilon)). \]

Using the fact that, given \( X(k+1) = z, X(1), \ldots, X(k) \) 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

\[ E[E(R_n | X(k+1))I(X(k+1) > 1)] \leq c(n\varepsilon)E[I(X(k+1) > 1)/X(k+1)] \leq c\varepsilon k/n \rightarrow 0. \]

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

\[ E(S_n | X(k+1)) = \int_{X(k+1)}^\infty \int_{x(1-\varepsilon)}^{x(1+\varepsilon)} w_\rho(x, y)dydx/ \int_{X(k+1)}^\infty h(x)dx \]

\[ = \int_{X(k+1)}^\infty h(x)P(Y \in [x(1-\varepsilon), x(1+\varepsilon)]|X) \]

\[ = xdx/ \int_{X(k+1)}^\infty h(x)dx. \]

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

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

for any \( \delta(x) \) with satisfying \( \delta(x) \rightarrow 0 \) and \( x\delta(x) \rightarrow \infty \). Hence, for \( X(k+1) > u_n \) we have

\[ E(S_n | X(k+1)) \sim \rho \]

and, similarly,

\[ \text{var}(S_n | X(k)) \leq (1/k) E(S_n | X(k+1)) \sim \rho/k. \]

It is then straightforward to conclude from these that \( S_n \xrightarrow{p} \rho. \) \( \square \)
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 A.1. (i) Let $A$ and $B$ be $m \times n$ matrices with singular values $\lambda_i$ and $\eta_i$ (both in descending order) respectively. Then,

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

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

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

Let the $\lambda_i$ and $\eta_i$ be the singular values (both in descending order) of $T_\phi$ and $T_\psi$, respectively. Then,

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

\[\Box\]

References


