Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conditional Estimation of Multivariate Dependence Using Projection-Based Bergsma–Dassios τ∗ and Nonparametric Hotelling's T2

STHITADHI DAS

sthdas999@gmail.com

Brainware University https://orcid.org/0009-0001-2484-9530

Research Article

Keywords: Bergsma–Dassios τ∗, multivariate dependence, nonparametric Hotelling statistic, U-statistic, Rao–Blackwell theorem, UMVUE, Stiefel manifold

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11631026/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Conditional Estimation of Multivariate Dependence Using Projection-Based Bergsma–Dassios τ^* and Nonparametric Hotelling's T^2

Sthitadhi Das*
Department of Mathematics
Brainware University, India

Abstract

This study develops a conditional estimation framework for a projection-based multivariate extension of the Bergsma–Dassios τ^* dependence measure. The proposed approach combines a fourth-order U-statistic estimator of multivariate dependence with a nonparametric Hotelling-type quadratic statistic. First, the unbiasedness and variance properties of the projection-based estimator are established. Subsequently, conditional estimation is investigated through the Rao–Blackwell principle. Under an unrestricted multivariate normal model with unknown mean vector and covariance matrix, the sample mean and sample scatter matrix are shown to constitute a complete sufficient statistic. An explicit conditional representation of the Rao–Blackwell estimator is then obtained through a uniform distribution on an appropriate Stiefel manifold. The Lehmann–Scheffé theorem establishes the resulting estimator as the unique uniformly minimum variance unbiased estimator (UMVUE) of the specified multivariate dependence parameter. Variance decomposition results are also derived to compare the original U-statistic, the estimator conditioned on the nonparametric Hotelling statistic, and the fully Rao–Blackwellized estimator.

Keywords: Bergsma–Dassios τ^*; multivariate dependence; nonparametric Hotelling statistic; U-statistic; Rao–Blackwell theorem; UMVUE; Stiefel manifold.

1 Introduction

The characterization and estimation of dependence between random vectors constitute fundamental problems in nonparametric statistics. Classical measures of association, such as Pearson's correlation coefficient and covariance, primarily quantify linear dependence and may fail to identify nonlinear relationships. This limitation has motivated the development of distribution-free dependence measures capable of detecting a broader class of relationships without imposing restrictive parametric assumptions. Among the established approaches, rank-based and concordance-based measures have received considerable attention because they exploit the relative ordering of observations and are generally less sensitive to marginal transformations than moment-based measures.

*Corresponding author: sthdas999@gmail.com

Several important developments have extended the study of dependence beyond classical correlation. Székely, Rizzo, and Bakirov [4] introduced distance covariance and distance correlation, which provide measures of dependence between random vectors of arbitrary dimensions and, under suitable moment conditions, characterize independence through the vanishing of distance covariance. In a different direction, Bergsma and Dassios [1] introduced the coefficient τ^*, a nonnegative measure of dependence between two univariate random variables based on a sign covariance involving four observations. Unlike conventional rank correlations that primarily characterize monotone association, τ^* is zero if and only if the two random variables are independent. Its four-observation structure also makes it particularly suitable for representation through U-statistics.

The extension of τ^* to multivariate random vectors presents an important methodological question. Although distance-based measures provide established multivariate dependence frameworks, a projection-based construction offers an alternative that retains the sign-kernel structure of the original τ^* coefficient. In particular, projecting two random vectors onto one-dimensional directions allows the univariate dependence measure to be evaluated over a collection of scalar projections. Averaging over the corresponding unit spheres then produces a scalar measure of multivariate dependence.

Motivated by this idea, the present study introduces and investigates a projection-based multivariate extension of the Bergsma–Dassios dependence measure. The proposed construction is defined by integrating the product of the two projected sign kernels over the unit spheres associated with the respective vector dimensions. The resulting population parameter is denoted by $\tau^*_{p,q}$. It is important to emphasize that this construction is a particular multivariate extension proposed in this work; its properties do not follow automatically from those of the original univariate coefficient or from other established multivariate dependence measures.

The four-observation structure of the proposed kernel naturally leads to an unbiased estimator based on a fourth-order U-statistic. The general theory of U-statistics, developed by Hoeffding [2], provides a systematic framework for establishing unbiasedness, deriving variance expressions, and studying asymptotic properties. In the present setting, the boundedness and symmetry of the projection-based kernel ensure that the corresponding U-statistic is well-defined and square-integrable. Its variance can be expressed through the Hoeffding decomposition, which separates the contributions of projections of different orders and provides a basis for analyzing the estimator's sampling variability.

Beyond constructing an unbiased estimator, an important objective is to investigate whether its variance can be reduced by conditioning on an appropriate statistic. Let

$$A = T^2_{\mathrm{NP}}$$

denote a specified nonparametric Hotelling-type statistic, and let

$$B = \widehat{\tau}^*_{p,q}$$

denote the fourth-order U-statistic estimator of $\tau^*_{p,q}$. We first consider the conditional quantity

$$B_A = \mathbb{E}(B \mid A).$$

By the law of iterated expectations, B_A remains unbiased whenever B is unbiased and integrable. Furthermore, the law of total variance gives

$$\mathrm{Var}(B) = \mathrm{Var}\{\mathbb{E}(B \mid A)\} + \mathbb{E}\{\mathrm{Var}(B \mid A)\},$$

which implies that conditioning cannot increase variance. Nevertheless, this variance reduction

identity alone does not establish that B_A is an implementable estimator: if the conditional distribution of B given A depends on unknown parameters, then $\mathbb{E}(B \mid A)$ may also depend on those parameters. Thus, the distinction between formal conditional estimation and parameter-free estimation is essential.

To obtain an explicit Rao–Blackwell estimator and investigate its optimality, we subsequently consider the unrestricted multivariate normal model

$$\mathbf{Z}_i \overset{\text{iid}}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad i = 1, \ldots, n,$$

where $\boldsymbol{\mu}$ and the positive-definite covariance matrix $\boldsymbol{\Sigma}$ are unknown. Under this model, the sample mean and sample scatter matrix jointly form a complete sufficient statistic for the unknown mean and covariance parameters. Writing

$$S = (\overline{\mathbf{Z}}, \mathbf{C}),$$

where

$$\overline{\mathbf{Z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i, \qquad \mathbf{C} = \sum_{i=1}^{n} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^T,$$

we define the Rao–Blackwellized estimator by

$$B_{\text{RB}} = \mathbb{E}(B \mid S).$$

The Rao–Blackwell theorem establishes that B_{RB} has variance no greater than that of the original unbiased estimator B. Moreover, the Lehmann–Scheffé theorem implies that an unbiased estimator expressed as a function of a complete sufficient statistic is the unique uniformly minimum variance unbiased estimator (UMVUE) of its target parameter [3]. Accordingly, under the specified multivariate normal model, the conditional estimator B_{RB} provides a candidate UMVUE of the projection-based dependence parameter $\tau^*_{p,q}$.

A further contribution of this work is an explicit representation of the conditional distribution required to evaluate B_{RB}. Conditional on the sample mean and scatter matrix, the centered sample configuration can be represented through a uniformly distributed orientation on a constrained Stiefel manifold. This representation yields a parameter-free integral expression for the Rao–Blackwell estimator and provides a computational approach based on Monte Carlo integration. It also clarifies the relationship between conditioning on the nonparametric Hotelling statistic and conditioning on the full complete sufficient statistic.

The principal objectives of this study are:

(i) to formulate a projection-based multivariate extension of the Bergsma–Dassios τ^* dependence measure and establish its fundamental kernel properties;

(ii) to construct an unbiased fourth-order U-statistic estimator and derive its variance through the Hoeffding decomposition;

(iii) to investigate conditional estimation based on a nonparametric Hotelling-type statistic and establish the associated variance reduction properties;

(iv) to derive an explicit conditional representation of the estimator given the complete sufficient statistic under multivariate normality;

(v) to establish the UMVUE property of the resulting Rao–Blackwell estimator under the specified Gaussian model; and

(vi) to develop variance decompositions and efficiency comparisons among the original U-statistic, the estimator conditioned on the nonparametric Hotelling statistic, and the fully Rao–Blackwellized estimator.

The remainder of the paper is organized as follows. Section 2 introduces the projection-based multivariate dependence measure and its U-statistic estimator. Section 3 develops the multivariate normal model and establishes the complete sufficiency of the sample mean and scatter matrix. Section 4 investigates conditional estimation and variance reduction. Section 5 derives the conditional Stiefel representation, followed by the explicit Rao–Blackwell estimator and its UMVUE characterization. The subsequent sections develop variance decompositions, computational procedures, and concluding remarks.

2 Projection-Based Multivariate Bergsma–Dassios Dependence Measure

2.1 Sampling Framework

Let

$$\{(\mathbf{X}_i, \mathbf{Y}_i) : i = 1, \dots, n\}$$

be an independent and identically distributed sample from the joint distribution of random vectors

$$\mathbf{X} \in \mathbb{R}^p, \qquad \mathbf{Y} \in \mathbb{R}^q.$$

For each observation, define the concatenated vector

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} \in \mathbb{R}^d, \qquad d = p + q.$$

The corresponding population vector is

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}.$$

The objective is to construct a scalar measure of dependence between $\mathbf{X}$ and $\mathbf{Y}$ by combining the univariate Bergsma–Dassios sign kernel with random one-dimensional projections. The original Bergsma–Dassios coefficient is based on a four-observation sign covariance and provides a distribution-free measure of dependence between two univariate random variables [1]. The construction below defines the particular projection-based multivariate extension studied in this paper.

2.2 Univariate Bergsma–Dassios Sign Kernel

For four real-valued observations z_1, z_2, z_3, z_4, define the ordered sign kernel

$$a(z_1, z_2, z_3, z_4) = \operatorname{sign}(|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|),$$

where

$$\operatorname{sign}(t) = \begin{cases} -1, & t < 0, \\ 0, & t = 0, \\ 1, & t > 0. \end{cases}$$

The kernel a is bounded by one in absolute value. However, it is not invariant under arbitrary permutations of its four arguments: the displayed expression distinguishes particular pairings of the observations. Consequently, symmetrizing this univariate kernel separately over all permutations does not provide a suitable nonzero symmetric kernel. Indeed, the terms corresponding to reversed pairings cancel in the full permutation average.

To obtain a symmetric kernel for the joint dependence problem, we instead apply the same permutation to the projected $\mathbf{X}$- and $\mathbf{Y}$-observations and symmetrize their product. This preserves the relative pairing structure between the two projected samples.

2.3 Projection-Based Multivariate Kernel

Let

$$\mathbb{S}^{p-1} = \{\boldsymbol{\alpha} \in \mathbb{R}^p : \|\boldsymbol{\alpha}\| = 1\}, \qquad \mathbb{S}^{q-1} = \{\boldsymbol{\beta} \in \mathbb{R}^q : \|\boldsymbol{\beta}\| = 1\}$$

denote the unit spheres in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Let λ_p and λ_q be the corresponding uniform probability measures on these spheres.

For four observations

$$\mathbf{z}_r = \begin{pmatrix} \mathbf{x}_r \\ \mathbf{y}_r \end{pmatrix}, \qquad r = 1, 2, 3, 4,$$

and directions $\boldsymbol{\alpha} \in \mathbb{S}^{p-1}$ and $\boldsymbol{\beta} \in \mathbb{S}^{q-1}$, define the projected sign kernels

$$\begin{aligned} a_{\boldsymbol{\alpha}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) &= a\big(\boldsymbol{\alpha}^T\mathbf{x}_1, \boldsymbol{\alpha}^T\mathbf{x}_2, \boldsymbol{\alpha}^T\mathbf{x}_3, \boldsymbol{\alpha}^T\mathbf{x}_4\big)\,, \\ a_{\boldsymbol{\beta}}(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4) &= a\big(\boldsymbol{\beta}^T\mathbf{y}_1, \boldsymbol{\beta}^T\mathbf{y}_2, \boldsymbol{\beta}^T\mathbf{y}_3, \boldsymbol{\beta}^T\mathbf{y}_4\big)\,. \end{aligned}$$

Let $\mathfrak{S}_4$ denote the permutation group on $\{1, 2, 3, 4\}$. Define the jointly symmetrized projection kernel by

$$\begin{aligned} g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) = \frac{1}{4!} \sum_{\pi \in \mathfrak{S}_4} & a\big(\boldsymbol{\alpha}^T\mathbf{x}_{\pi(1)}, \boldsymbol{\alpha}^T\mathbf{x}_{\pi(2)}, \boldsymbol{\alpha}^T\mathbf{x}_{\pi(3)}, \boldsymbol{\alpha}^T\mathbf{x}_{\pi(4)}\big) \\ & \times a\big(\boldsymbol{\beta}^T\mathbf{y}_{\pi(1)}, \boldsymbol{\beta}^T\mathbf{y}_{\pi(2)}, \boldsymbol{\beta}^T\mathbf{y}_{\pi(3)}, \boldsymbol{\beta}^T\mathbf{y}_{\pi(4)}\big)\,. \end{aligned}$$

The same permutation π is used in both factors. This is essential: it preserves the correspondence between the four projected $\mathbf{X}$- and $\mathbf{Y}$-observations. Moreover, a common permutation of the four joint observations merely reorders the summands, so $g_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is symmetric in its four joint arguments.

The projection-based multivariate kernel is defined by averaging this jointly symmetrized kernel over the two unit spheres:

$$\boxed{h(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) = \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^{q-1}} g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4)\, d\lambda_q(\boldsymbol{\beta})\, d\lambda_p(\boldsymbol{\alpha}).}$$

Equivalently, the kernel can be written explicitly as

$$\begin{aligned} h(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) = \frac{1}{4!} \sum_{\pi \in \mathfrak{S}_4} & \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^{q-1}} \\ & a\big(\boldsymbol{\alpha}^T\mathbf{x}_{\pi(1)}, \boldsymbol{\alpha}^T\mathbf{x}_{\pi(2)}, \boldsymbol{\alpha}^T\mathbf{x}_{\pi(3)}, \boldsymbol{\alpha}^T\mathbf{x}_{\pi(4)}\big) \\ & \times a\big(\boldsymbol{\beta}^T\mathbf{y}_{\pi(1)}, \boldsymbol{\beta}^T\mathbf{y}_{\pi(2)}, \boldsymbol{\beta}^T\mathbf{y}_{\pi(3)}, \boldsymbol{\beta}^T\mathbf{y}_{\pi(4)}\big) \\ & d\lambda_q(\boldsymbol{\beta})\, d\lambda_p(\boldsymbol{\alpha}). \end{aligned}$$

Since each sign kernel takes values in $[-1, 1]$,

$$|g_{\boldsymbol{\alpha},\boldsymbol{\beta}}| \le 1 \qquad \text{and} \qquad |h| \le 1.$$

Thus h is bounded and integrable. By construction, h is also symmetric in its four joint arguments.

Definition 2.1. The projection-based multivariate Bergsma–Dassios dependence parameter is

$$\boxed{\theta = \tau^*_{p,q} = \mathbb{E}[h(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4)]\,,}$$

where $\mathbf{Z}_1, \ldots, \mathbf{Z}_4$ are independent copies of $\mathbf{Z}$.

This definition specifies the multivariate extension considered in the present work. Its properties, including its behavior under independence and its dependence on the marginal distributions, require separate analysis and are not assumed to follow automatically from the corresponding properties of the original univariate τ^* coefficient.

2.4 U-Statistic Estimator

For $n \ge 4$, define

$$\boxed{B = \widehat{\tau}^*_{p,q} = \binom{n}{4}^{-1} \sum_{1 \le i < j < k < l \le n} h(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \mathbf{Z}_l).}$$

Because h is symmetric and has order four, B is a fourth-order U-statistic with kernel h, in the sense of the general theory of U-statistics [2].

Theorem 2.2 (Unbiasedness)**.** *Suppose that $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ are iid and that $n \ge 4$. Then*

$$\mathbb{E}(B) = \tau^*_{p,q}.$$

Proof. By the definition of B and linearity of expectation,

$$\mathbb{E}(B) = \binom{n}{4}^{-1} \sum_{1 \le i < j < k < l \le n} \mathbb{E}[h(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \mathbf{Z}_l)]\,.$$

For every index tuple $i < j < k < l$, the four observations $\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \mathbf{Z}_l$ are independent and identically distributed. Therefore,

$$\begin{aligned} &\mathbb{E}[h(\mathbf{Z}_i, \mathbf{Z}_j, \mathbf{Z}_k, \mathbf{Z}_l)] \\ &\qquad = \mathbb{E}[h(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4)] = \tau^*_{p,q}. \end{aligned}$$

There are exactly $\binom{n}{4}$ terms in the sum. Consequently,

$$\mathbb{E}(B) = \binom{n}{4}^{-1} \binom{n}{4} \tau^*_{p,q} = \tau^*_{p,q},$$

which proves the result. □

The boundedness of h also guarantees that $\mathbb{E}(B^2) < \infty$. Hence the estimator has finite variance, and its variance can be studied using the Hoeffding decomposition for U-statistics [2].

3 Distributional Model and Sufficient Statistics

For the explicit development of the Rao–Blackwell estimator and its uniformly minimum variance unbiased estimator (UMVUE) property, we consider the unrestricted multivariate normal model

$$\boxed{\mathbf{Z}_i \overset{\text{iid}}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad i = 1, \ldots, n,}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ is an unknown mean vector and $\boldsymbol{\Sigma}$ is an unknown positive-definite $d \times d$ covariance matrix, with $d = p + q$.

Recall that

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix}, \qquad \mathbf{X}_i \in \mathbb{R}^p, \qquad \mathbf{Y}_i \in \mathbb{R}^q.$$

Accordingly, the mean vector and covariance matrix can be partitioned as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{pmatrix},$$

where $\boldsymbol{\mu}_X \in \mathbb{R}^p$, $\boldsymbol{\mu}_Y \in \mathbb{R}^q$, $\boldsymbol{\Sigma}_{XX}$ and $\boldsymbol{\Sigma}_{YY}$ are the marginal covariance matrices, and $\boldsymbol{\Sigma}_{XY} = \boldsymbol{\Sigma}_{YX}^T$ is the cross-covariance matrix.

The positive definiteness of $\boldsymbol{\Sigma}$ guarantees that both marginal covariance matrices are positive definite and that the projected variances considered below are strictly positive for all nonzero projection directions.

3.1 Population Parameter under Multivariate Normality

Consider fixed projection directions

$$\boldsymbol{\alpha} \in \mathbb{S}^{p-1}, \qquad \boldsymbol{\beta} \in \mathbb{S}^{q-1},$$

and define the corresponding scalar projections

$$U = \boldsymbol{\alpha}^T \mathbf{X}, \qquad V = \boldsymbol{\beta}^T \mathbf{Y}.$$

Since $(\mathbf{X}^T, \mathbf{Y}^T)^T$ follows a multivariate normal distribution, the projected pair (U, V) follows a bivariate normal distribution.

The marginal means of these projections are

$$\mathbb{E}(U) = \boldsymbol{\alpha}^T \boldsymbol{\mu}_X, \qquad \mathbb{E}(V) = \boldsymbol{\beta}^T \boldsymbol{\mu}_Y,$$

while their variances are

$$\operatorname{Var}(U) = \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{XX} \boldsymbol{\alpha}$$

and

$$\operatorname{Var}(V) = \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{YY} \boldsymbol{\beta}.$$

Furthermore, their covariance is

$$\operatorname{Cov}(U, V) = \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{XY} \boldsymbol{\beta}.$$

Consequently, the correlation between the two projected random variables is

$$\boxed{\rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{XY} \boldsymbol{\beta}}{\sqrt{(\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{XX} \boldsymbol{\alpha})\,(\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{YY} \boldsymbol{\beta})}}.}$$

The positive definiteness of $\boldsymbol{\Sigma}$ ensures that the denominator is strictly positive. In particular,

$$-1 \leq \rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq 1.$$

Let $\tau_{\mathrm{G}}^*(\rho)$ denote the population Bergsma–Dassios coefficient for a bivariate normal pair with standard normal marginal distributions and correlation coefficient ρ, using the same sign-kernel definition and normalization as in the preceding section. The coefficient is invariant under strictly increasing affine transformations of either marginal variable. Therefore, the population coefficient for the projected pair (U, V) depends on its distribution only through the correlation parameter $\rho(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

For each fixed pair of projection directions, we therefore have

$$\begin{aligned}\mathbb{E}\left[a(U_1, U_2, U_3, U_4)a(V_1, V_2, V_3, V_4)\right] \\ = \tau_{\mathrm{G}}^*\left(\rho(\boldsymbol{\alpha}, \boldsymbol{\beta})\right),\end{aligned}$$

where (U_i, V_i), $i = 1, \ldots, 4$, are independent copies of (U, V).

By the definition of the projection-based kernel, linearity of expectation, and Fubini's theorem, the multivariate dependence parameter can consequently be expressed as

$$\boxed{\begin{aligned}\tau_{p,q}^* = \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^{q-1}} \tau_{\mathrm{G}}^*\left(\rho(\boldsymbol{\alpha}, \boldsymbol{\beta})\right) \\ d\lambda_q(\boldsymbol{\beta}) d\lambda_p(\boldsymbol{\alpha}).\end{aligned}}$$

Here, the use of the same permutation in the two projected sign kernels ensures that the expectation of the jointly symmetrized product equals the corresponding univariate Bergsma–Dassios expectation for each fixed pair of directions.

This representation establishes an important simplification under the multivariate normal model. The projection-based dependence parameter is determined entirely by the covariance structure of the joint distribution. In particular,

$$\boxed{\tau_{p,q}^* = \Psi(\boldsymbol{\Sigma}),}$$

for an appropriate function Ψ determined by the spherical integration. Thus, the parameter does not depend on the unknown mean vector $\boldsymbol{\mu}$.

It is worth noting that the above representation is specific to the multivariate normal model. For general joint distributions, the projected Bergsma–Dassios coefficient need not be determined solely by the projected correlation.

3.2 Likelihood Function and Sufficient Statistic

Under the assumed multivariate normal model, the joint density of the sample is

$$\begin{aligned}f(\mathbf{z}_1, \ldots, \mathbf{z}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-nd/2} |\boldsymbol{\Sigma}|^{-n/2} \\ \times \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu})\right\}.\end{aligned}$$

Define the two statistics

$$\boxed{S_1 = \sum_{i=1}^{n} \mathbf{Z}_i, \qquad S_2 = \sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i^T.}$$

Expanding the quadratic expression in the likelihood gives

$$\begin{aligned}
&\sum_{i=1}^{n}(\mathbf{Z}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Z}_i-\boldsymbol{\mu}) \\
&\quad=\sum_{i=1}^{n}\mathbf{Z}_i^T\boldsymbol{\Sigma}^{-1}\mathbf{Z}_i-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\mathbf{Z}_i+n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
&\quad=\operatorname{tr}(\boldsymbol{\Sigma}^{-1}S_2)-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}S_1+n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}.
\end{aligned}$$

Hence, the likelihood can be expressed as

$$\begin{aligned}
L(\boldsymbol{\mu},\boldsymbol{\Sigma})\propto{}&|\boldsymbol{\Sigma}|^{-n/2} \\
&\times\exp\left\{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}S_2)+\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}S_1-\frac{n}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\}.
\end{aligned}$$

The sample enters the likelihood only through (S_1, S_2). Therefore, by the factorization theorem, the statistic

$$\boxed{S=(S_1,S_2)}$$

is sufficient for $(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

3.3 Exponential-Family Representation and Completeness

The joint density can also be written in the canonical exponential-family form

$$f(\mathbf{z}_1,\ldots,\mathbf{z}_n;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)=\exp\left\{\boldsymbol{\eta}_1^TS_1+\operatorname{tr}(\boldsymbol{\eta}_2S_2)-A(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)\right\},$$

where the natural parameters are

$$\boxed{\boldsymbol{\eta}_1=\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},\qquad\boldsymbol{\eta}_2=-\frac{1}{2}\boldsymbol{\Sigma}^{-1}.}$$

The corresponding log-partition function is

$$\begin{aligned}
A(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2)=&\frac{nd}{2}\log(2\pi)-\frac{n}{2}\log|\boldsymbol{\Sigma}| \\
&+\frac{n}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},
\end{aligned}$$

with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ understood as functions of the natural parameters.

Since $\boldsymbol{\Sigma}$ is positive definite, $\boldsymbol{\eta}_2$ is negative definite. Moreover, $\boldsymbol{\eta}_1$ can take any value in $\mathbb{R}^d$. Thus, the natural parameter space is

$$\mathcal{H}=\left\{(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2):\boldsymbol{\eta}_1\in\mathbb{R}^d,\ \boldsymbol{\eta}_2=\boldsymbol{\eta}_2^T<0\right\},$$

where $\boldsymbol{\eta}_2<0$ means that $\boldsymbol{\eta}_2$ is negative definite.

This is an open subset of the finite-dimensional natural parameter space

$$\mathbb{R}^d\times\mathbb{S}^d,$$

where $\mathbb{S}^d$ denotes the space of real symmetric $d\times d$ matrices. The exponential family is therefore a full-rank exponential family with an open natural parameter space.

The completeness of the natural sufficient statistic then follows from the completeness

theorem for full-rank exponential families with open natural parameter spaces; see, for example, Lehmann and Casella (1998).

Theorem 3.1 (Completeness and sufficiency)**.** *Suppose that*

$$\mathbf{Z}_i \overset{\text{iid}}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad i = 1, \ldots, n,$$

where $\boldsymbol{\mu}$ *is unrestricted in* $\mathbb{R}^d$ *and* $\boldsymbol{\Sigma}$ *is unrestricted over the set of positive-definite symmetric* $d \times d$ *matrices. Then*

$$\boxed{S = \left(\sum_{i=1}^{n} \mathbf{Z}_i, \; \sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i^T \right)}$$

is a complete and sufficient statistic for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Proof. The likelihood factorization established above shows that $S = (S_1, S_2)$ is sufficient.

Furthermore, the multivariate normal family with unknown mean vector and unrestricted positive-definite covariance matrix is a full-rank exponential family. Its natural parameter space is open in the corresponding finite-dimensional natural parameter space. Consequently, the completeness theorem for full-rank exponential families applies, establishing that the natural sufficient statistic (S_1, S_2) is complete.

Thus, S is both complete and sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. □

3.4 Equivalent Representation in Terms of the Sample Mean and Scatter Matrix

For subsequent developments, it is convenient to express the complete sufficient statistic in terms of the sample mean and the sample scatter matrix.

Define

$$\boxed{\overline{\mathbf{Z}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Z}_i}$$

and

$$\boxed{\mathbf{C} = \sum_{i=1}^{n} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^T.}$$

Expanding the scatter matrix gives

$$\begin{aligned} \mathbf{C} &= \sum_{i=1}^{n} \left(\mathbf{Z}_i \mathbf{Z}_i^T - \mathbf{Z}_i \overline{\mathbf{Z}}^T - \overline{\mathbf{Z}} \mathbf{Z}_i^T + \overline{\mathbf{Z}} \overline{\mathbf{Z}}^T \right) \\ &= \sum_{i=1}^{n} \mathbf{Z}_i \mathbf{Z}_i^T - n \overline{\mathbf{Z}} \overline{\mathbf{Z}}^T. \end{aligned}$$

Therefore,

$$S_1 = n\overline{\mathbf{Z}}, \qquad S_2 = \mathbf{C} + n\overline{\mathbf{Z}}\overline{\mathbf{Z}}^T.$$

Conversely,

$$\overline{\mathbf{Z}} = \frac{S_1}{n}, \qquad \mathbf{C} = S_2 - \frac{1}{n} S_1 S_1^T.$$

Hence, (S_1, S_2) and $(\overline{\mathbf{Z}}, \mathbf{C})$ are one-to-one transformations of each other.

It follows that the statistic

$$\boxed{S = (\overline{\mathbf{Z}}, \mathbf{C})}$$

is also complete and sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

This representation will be used in the subsequent Rao–Blackwell development. In particular, the conditional distribution of the sample configuration given $(\overline{\mathbf{Z}}, \mathbf{C})$ will provide the basis for constructing an explicit conditional expectation of the original U-statistic.

4 Conditional Estimation

4.1 Conditional Estimator Based on the Nonparametric Hotelling Statistic

Let

$$A = T_{\mathrm{NP}}^2$$

denote a specified nonparametric Hotelling-type statistic constructed from the observed sample, and let

$$B = \widehat{\tau}_{p,q}^{*}$$

denote the unbiased fourth-order U-statistic estimator of the projection-based multivariate dependence parameter

$$\theta = \tau_{p,q}^{*}.$$

Our first objective is to investigate whether the estimator B can be improved by conditioning on the statistic A. For this purpose, define

$$\boxed{B_A = \mathbb{E}(B \mid A).}$$

Here, $\mathbb{E}(B \mid A)$ denotes the conditional expectation of B with respect to the σ-field generated by A, denoted by $\sigma(A)$.

The conditional estimator B_A is the best mean-square predictor of B among all square-integrable random variables measurable with respect to $\sigma(A)$. Thus, it represents the component of B that can be explained using the information contained in A.

Theorem 4.1 (Conditional unbiasedness)**.** *Suppose that B is an unbiased estimator of θ and that*

$$\mathbb{E}|B| < \infty.$$

Then the conditional estimator

$$B_A = \mathbb{E}(B \mid A)$$

is also unbiased for θ, that is,

$$\boxed{\mathbb{E}(B_A) = \theta.}$$

Proof. By the definition of conditional expectation,

$$B_A = \mathbb{E}(B \mid A).$$

Taking expectations on both sides and applying the law of iterated expectations, we obtain

$$\begin{aligned} \mathbb{E}(B_A) &= \mathbb{E}\{\mathbb{E}(B \mid A)\} \\ &= \mathbb{E}(B). \end{aligned}$$

Since B is unbiased for θ,

$$\mathbb{E}(B) = \theta.$$

Therefore,

$$\boxed{\mathbb{E}(B_A) = \theta.}$$

This establishes the unbiasedness of the conditional estimator. □

Remark 4.2. The unbiasedness result is distribution-free in the sense that it follows directly from the law of iterated expectations and does not require A to be sufficient.

However, unbiasedness alone does not guarantee that B_A is an operationally available estimator. If the conditional distribution of B given A depends on unknown population parameters, then the conditional expectation $\mathbb{E}(B \mid A)$ may also involve unknown parameters. In such a situation, B_A is a well-defined conditional random variable but may not be directly computable from the observed sample.

Consequently, a parameter-free conditional distribution is required to obtain an implementable conditional estimator. This distinction motivates the subsequent development based on the complete sufficient statistic under the multivariate normal model.

4.2 Conditional Expectation and Conditional Variance

The conditional expectation and conditional variance describe the behavior of one random variable when information about another random variable is available.

Suppose initially that the pair (A, B) admits a joint probability density function $f_{A,B}(a, b)$, with marginal densities $f_A(a)$ and $f_B(b)$. Whenever $f_A(a) > 0$, the conditional density of B given $A = a$ is

$$\boxed{f_{B|A}(b \mid a) = \frac{f_{A,B}(a, b)}{f_A(a)}.}$$

Therefore, the conditional expectation of B given $A = a$ is

$$\boxed{\mathbb{E}(B \mid A = a) = \frac{\displaystyle\int_{-\infty}^{\infty} b f_{A,B}(a, b)\, db}{f_A(a)}.}$$

Similarly, the conditional second moment is

$$\mathbb{E}(B^2 \mid A = a) = \frac{\displaystyle\int_{-\infty}^{\infty} b^2 f_{A,B}(a, b)\, db}{f_A(a)}.$$

Using the identity

$$\mathrm{Var}(B \mid A = a) = \mathbb{E}(B^2 \mid A = a) - \{\mathbb{E}(B \mid A = a)\}^2,$$

we obtain

$$\boxed{\mathrm{Var}(B \mid A = a) = \frac{\displaystyle\int_{-\infty}^{\infty} b^2 f_{A,B}(a, b)\, db}{f_A(a)} - \left[\frac{\displaystyle\int_{-\infty}^{\infty} b f_{A,B}(a, b)\, db}{f_A(a)}\right]^2.}$$

In the reverse direction, whenever $f_B(b) > 0$, the conditional density of A given $B = b$ is

$$f_{A|B}(a \mid b) = \frac{f_{A,B}(a,b)}{f_B(b)}.$$

Consequently,

$$\boxed{\mathbb{E}(A \mid B = b) = \frac{\displaystyle\int_{-\infty}^{\infty} a f_{A,B}(a,b)\, da}{f_B(b)}.}$$

The corresponding conditional second moment is

$$\mathbb{E}(A^2 \mid B = b) = \frac{\displaystyle\int_{-\infty}^{\infty} a^2 f_{A,B}(a,b)\, da}{f_B(b)}.$$

Hence,

$$\boxed{\begin{aligned}\mathrm{Var}(A \mid B = b) = &\frac{\displaystyle\int_{-\infty}^{\infty} a^2 f_{A,B}(a,b)\, da}{f_B(b)} \\ &- \left[\frac{\displaystyle\int_{-\infty}^{\infty} a f_{A,B}(a,b)\, da}{f_B(b)}\right]^2.\end{aligned}}$$

These expressions provide the density-based representations of the conditional moments. If the joint distribution does not admit a density, the corresponding conditional expectations and variances are understood through regular conditional distributions.

4.3 Variance Reduction Through Conditioning

An important property of conditional expectation is that it cannot increase the variance of a square-integrable random variable. This property provides the fundamental theoretical justification for conditioning-based estimator improvement.

Theorem 4.3 (Conditional variance decomposition)**.** *Let B be a square-integrable random variable and let A be any statistic. Then*

$$\boxed{\mathrm{Var}(B) = \mathrm{Var}\{\mathbb{E}(B \mid A)\} + \mathbb{E}\{\mathrm{Var}(B \mid A)\}.}$$

Consequently,

$$\boxed{\mathrm{Var}(B_A) \le \mathrm{Var}(B).}$$

Proof. By the definition of conditional variance,

$$\mathrm{Var}(B \mid A) = \mathbb{E}(B^2 \mid A) - \{\mathbb{E}(B \mid A)\}^2.$$

Rearranging gives

$$\mathbb{E}(B^2 \mid A) = \mathrm{Var}(B \mid A) + \{\mathbb{E}(B \mid A)\}^2.$$

Taking expectations on both sides, we obtain

$$\mathbb{E}(B^2) = \mathbb{E}\{\mathrm{Var}(B \mid A)\} + \mathbb{E}[\{\mathbb{E}(B \mid A)\}^2].$$

Furthermore, the law of iterated expectations implies that

$$\mathbb{E}\{\mathbb{E}(B \mid A)\} = \mathbb{E}(B).$$

Therefore,

$$\begin{aligned}\mathrm{Var}(B) &= \mathbb{E}(B^2) - \{\mathbb{E}(B)\}^2 \\ &= \mathbb{E}\{\mathrm{Var}(B \mid A)\} \\ &\quad + \mathbb{E}[\{\mathbb{E}(B \mid A)\}^2] - \{\mathbb{E}(B)\}^2 \\ &= \mathbb{E}\{\mathrm{Var}(B \mid A)\} + \mathrm{Var}\{\mathbb{E}(B \mid A)\}.\end{aligned}$$

Since conditional variance is nonnegative,

$$\mathrm{Var}\{\mathbb{E}(B \mid A)\} \leq \mathrm{Var}(B).$$

The result follows. □

Remark 4.4 (Condition for equality). Equality in the conditional variance inequality holds if and only if

$$\mathrm{Var}(B \mid A) = 0 \qquad \text{almost surely.}$$

Equivalently, there exists a measurable function ϕ such that

$$B = \phi(A) \qquad \text{almost surely.}$$

Thus, conditioning produces a strict variance reduction whenever B retains nonzero conditional variation after observing A.

For the estimator considered in this study, the variance decomposition becomes

$$\boxed{\mathrm{Var}\left(\widehat{\tau}^*_{p,q}\right) = \mathrm{Var}\left\{\mathbb{E}\left(\widehat{\tau}^*_{p,q} \mid T^2_{\mathrm{NP}}\right)\right\} + \mathbb{E}\left\{\mathrm{Var}\left(\widehat{\tau}^*_{p,q} \mid T^2_{\mathrm{NP}}\right)\right\}.}$$

This identity establishes a theoretical variance reduction property of conditional expectation. However, the parameter-free nature of the resulting conditional estimator must be investigated separately.

5 Conditional Distribution Given the Complete Sufficient Statistic

We now develop an explicit conditional representation under the unrestricted multivariate normal model

$$\mathbf{Z}_i \overset{\text{iid}}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad i = 1, \ldots, n.$$

From the preceding section, the complete sufficient statistic for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$S = (\overline{\mathbf{Z}}, \mathbf{C}),$$

where

$$\overline{\mathbf{Z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i$$

and

$$\mathbf{C} = \sum_{i=1}^{n} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^T.$$

The objective of this section is to characterize the conditional distribution of the complete sample configuration given the observed values of $\overline{\mathbf{Z}}$ and $\mathbf{C}$.

5.1 Centered Sample Matrix

Define the centered sample matrix

$$\mathbf{Z}_c = \begin{pmatrix} (\mathbf{Z}_1 - \overline{\mathbf{Z}})^T \\ \vdots \\ (\mathbf{Z}_n - \overline{\mathbf{Z}})^T \end{pmatrix} \in \mathbb{R}^{n \times d}.$$

By construction,

$$\boxed{\mathbf{Z}_c^T \mathbf{Z}_c = \mathbf{C}}$$

and

$$\boxed{\mathbf{1}_n^T \mathbf{Z}_c = \mathbf{0}^T,}$$

where $\mathbf{1}_n$ denotes the n-dimensional vector of ones.

The first condition specifies the observed scatter matrix, whereas the second condition expresses the centering constraint.

Assume throughout this section that

$$n > d.$$

Under the multivariate normal model, the sample scatter matrix $\mathbf{C}$ is positive definite with probability one under this condition.

Let $\mathbf{C}^{1/2}$ denote the unique symmetric positive-definite square root of $\mathbf{C}$. Define

$$\boxed{\mathbf{Q} = \mathbf{Z}_c \mathbf{C}^{-1/2}.}$$

Then

$$\begin{aligned} \mathbf{Q}^T \mathbf{Q} &= \mathbf{C}^{-1/2} \mathbf{Z}_c^T \mathbf{Z}_c \mathbf{C}^{-1/2} \\ &= \mathbf{C}^{-1/2} \mathbf{C} \mathbf{C}^{-1/2} \\ &= \mathbf{I}_d. \end{aligned}$$

Also,

$$\mathbf{1}_n^T \mathbf{Q} = \mathbf{1}_n^T \mathbf{Z}_c \mathbf{C}^{-1/2} = \mathbf{0}^T.$$

Consequently,

$$\mathbf{Q} \in \mathcal{V}_{d,n-1},$$

where the constrained Stiefel manifold is defined by

$$\boxed{\mathcal{V}_{d,n-1} = \left\{ \mathbf{Q} \in \mathbb{R}^{n \times d} : \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_d,\ \mathbf{1}_n^T \mathbf{Q} = \mathbf{0}^T \right\}.}$$

Conversely, every matrix $\mathbf{Q}$ belonging to this manifold generates a centered sample matrix satisfying the observed scatter constraint through

$$\boxed{\mathbf{Z}_c = \mathbf{Q} \mathbf{C}^{1/2}.}$$

Therefore, the complete sample configuration can be represented as

$$\boxed{\mathbf{Z}_i(\mathbf{Q}) = \overline{\mathbf{Z}} + \mathbf{C}^{1/2} \mathbf{q}_i, \qquad i = 1, \ldots, n,}$$

where $\mathbf{q}_i$ denotes the transpose of the ith row of $\mathbf{Q}$.

Thus, once the sample mean and scatter matrix are fixed, the remaining uncertainty concerning the sample configuration is represented entirely by the orientation matrix $\mathbf{Q}$.

5.2 Conditional Stiefel Distribution

Lemma 5.1 (Conditional Stiefel distribution)**.** *Suppose that*

$$\mathbf{Z}_i \overset{\text{iid}}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ is unrestricted and $\boldsymbol{\Sigma}$ is positive definite. If $n > d$, then

$$\boxed{\mathbf{Q} \mid (\overline{\mathbf{Z}}, \mathbf{C}) \sim \nu_{d,n-1},}$$

where $\nu_{d,n-1}$ denotes the uniform invariant probability measure on the constrained Stiefel manifold $\mathcal{V}_{d,n-1}$.

In particular, this conditional distribution does not depend on either $\boldsymbol{\mu}$ or $\boldsymbol{\Sigma}$.

Proof. Let $\mathbf{H}$ be an $n \times (n-1)$ matrix satisfying

$$\mathbf{H}^T\mathbf{H} = \mathbf{I}_{n-1}, \qquad \mathbf{H}^T\mathbf{1}_n = \mathbf{0}.$$

Thus, the columns of $\mathbf{H}$ form an orthonormal basis for the subspace orthogonal to $\mathbf{1}_n$.

Since the columns of $\mathbf{Z}_c$ are centered, there exists an $(n-1) \times d$ matrix $\mathbf{R}$ such that

$$\mathbf{Z}_c = \mathbf{H}\mathbf{R}.$$

Moreover,

$$\mathbf{R}^T\mathbf{R} = \mathbf{Z}_c^T\mathbf{H}\mathbf{H}^T\mathbf{Z}_c = \mathbf{Z}_c^T\mathbf{Z}_c = \mathbf{C}.$$

Under the multivariate normal model, the centered sample matrix has a matrix-normal distribution. In particular, after projection onto the $(n-1)$-dimensional subspace orthogonal to $\mathbf{1}_n$, the matrix $\mathbf{R}$ has a matrix-normal distribution with zero mean and column covariance matrix $\boldsymbol{\Sigma}$. Its density is proportional to

$$|\boldsymbol{\Sigma}|^{-(n-1)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{R}^T\mathbf{R}\right)\right\}.$$

Using

$$\mathbf{R}^T\mathbf{R} = \mathbf{C},$$

we observe that the conditional distribution of the orientation of $\mathbf{R}$, given its Gram matrix $\mathbf{C}$, is invariant under left orthogonal transformations in the $(n-1)$-dimensional centered observation space.

The orthogonal group acts transitively on the set of matrices satisfying

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_d, \qquad \mathbf{1}_n^T\mathbf{Q} = \mathbf{0}^T.$$

Consequently, the unique invariant probability measure on this set is the uniform Stiefel measure.

Since

$$\mathbf{Q} = \mathbf{Z}_c\mathbf{C}^{-1/2},$$

the conditional distribution of $\mathbf{Q}$ given $(\overline{\mathbf{Z}}, \mathbf{C})$ is precisely this uniform invariant distribution.

Therefore,

$$\mathbf{Q} \mid (\overline{\mathbf{Z}}, \mathbf{C}) \sim \nu_{d,n-1},$$

and the conditional distribution is independent of the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. □

Remark 5.2. The parameter-free nature of the conditional Stiefel distribution is the central property required for the explicit Rao–Blackwell construction. It allows the conditional expectation to be evaluated using only the observed sample mean and scatter matrix, without requiring knowledge of the unknown population parameters.

6 Explicit Rao–Blackwell Estimator

Let

$$S = (\overline{\mathbf{Z}}, \mathbf{C})$$

denote the complete sufficient statistic under the unrestricted multivariate normal model.

The Rao–Blackwell estimator associated with the unbiased U-statistic $B = \widehat{\tau}^*_{p,q}$ is defined by

$$\boxed{B_{\mathrm{RB}} = \mathbb{E}(B \mid S).}$$

Since the conditional distribution of the sample configuration given S is uniform on the constrained Stiefel manifold, the conditional expectation admits an explicit integral representation.

6.1 Integral Representation

Using

$$\mathbf{Z}_i(\mathbf{Q}) = \overline{\mathbf{Z}} + \mathbf{C}^{1/2}\mathbf{q}_i,$$

the Rao–Blackwell estimator can be written as

$$\boxed{B_{\mathrm{RB}} = \int_{\mathcal{V}_{d,n-1}} \widehat{\tau}^*_{p,q}\left(\mathbf{Z}_1(\mathbf{Q}), \ldots, \mathbf{Z}_n(\mathbf{Q})\right) d\nu_{d,n-1}(\mathbf{Q}).} \tag{1}$$

Substituting the fourth-order U-statistic representation gives

$$B_{\mathrm{RB}} = \int_{\mathcal{V}_{d,n-1}} \binom{n}{4}^{-1} \sum_{1 \le i < j < k < l \le n} h\left(\mathbf{Z}_i(\mathbf{Q}), \mathbf{Z}_j(\mathbf{Q}), \mathbf{Z}_k(\mathbf{Q}), \mathbf{Z}_l(\mathbf{Q})\right) d\nu_{d,n-1}(\mathbf{Q}).$$

Since the summation contains finitely many terms, the integral and summation can be interchanged. Therefore,

$$\boxed{B_{\mathrm{RB}} = \binom{n}{4}^{-1} \sum_{1 \le i < j < k < l \le n} \int_{\mathcal{V}_{d,n-1}} h\big(\mathbf{Z}_i(\mathbf{Q}), \mathbf{Z}_j(\mathbf{Q}), \mathbf{Z}_k(\mathbf{Q}), \mathbf{Z}_l(\mathbf{Q})\big) d\nu_{d,n-1}(\mathbf{Q}).} \tag{2}$$

Equation (2) provides the desired explicit conditional-integral representation of the Rao–Blackwell estimator.

6.2 Properties of the Conditional Estimator

The boundedness of the original kernel implies that

$$|B| \le 1.$$

Consequently,

$$|B_{\mathrm{RB}}| = |\mathbb{E}(B \mid S)| \le \mathbb{E}(|B| \mid S) \le 1.$$

Thus, the Rao–Blackwell estimator is also bounded and has finite variance.

Furthermore, since the integration in (1) is performed with respect to a parameter-free probability measure, B_{RB} is a function of the observed sufficient statistic alone.

Remark 6.1 (Computational interpretation). The integral in (2) is an exact mathematical representation. In general, it need not admit a closed-form analytical expression.

A practical numerical approximation can be obtained by generating independent matrices

$$\mathbf{Q}^{(1)}, \dots, \mathbf{Q}^{(M)}$$

from the uniform invariant distribution on $\mathcal{V}_{d,n-1}$ and computing

$$\boxed{\widehat{B}_{\mathrm{RB},M} = \frac{1}{M} \sum_{m=1}^{M} \widehat{\tau}^{*}_{p,q} \left(\mathbf{Z}_1(\mathbf{Q}^{(m)}), \dots, \mathbf{Z}_n(\mathbf{Q}^{(m)}) \right).}$$

Conditional on the observed sufficient statistic, this Monte Carlo approximation is unbiased for B_{RB} and converges to the conditional integral as $M \to \infty$ under the usual Monte Carlo conditions.

7 UMVUE Characterization

We now establish the uniformly minimum variance unbiased estimator property of the proposed Rao–Blackwell estimator.

Theorem 7.1 (UMVUE of the multivariate Bergsma–Dassios dependence parameter)**.** *Suppose that the following conditions hold:*

(i) The observations $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ *are iid from the unrestricted multivariate normal family*

$$N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ *is unrestricted and* $\boldsymbol{\Sigma}$ *is positive definite.*

(ii) The sample size satisfies

$$n > d, \qquad n \ge 4.$$

(iii) The estimator

$$B = \widehat{\tau}^{*}_{p,q}$$

is the bounded unbiased fourth-order U-statistic defined in the preceding section.

(iv) The target parameter is

$$\theta = \tau^{*}_{p,q} = \mathbb{E}(B).$$

Then

$$\boxed{B_{\mathrm{RB}} = \mathbb{E}(B \mid S), \qquad S = (\overline{\mathbf{Z}}, \mathbf{C}),}$$

is the unique uniformly minimum variance unbiased estimator of θ under the specified multivariate normal model.

Proof. We establish the result by verifying unbiasedness, sufficiency, completeness, and the measurability of the conditional estimator.

First, since B is unbiased for θ,

$$\mathbb{E}(B) = \theta.$$

By the law of iterated expectations,

$$\begin{aligned} \mathbb{E}(B_{\mathrm{RB}}) &= \mathbb{E}\{\mathbb{E}(B \mid S)\} \\ &= \mathbb{E}(B) \\ &= \theta. \end{aligned}$$

Thus, B_{RB} is unbiased for θ.

Second, the statistic

$$S = (\overline{\mathbf{Z}}, \mathbf{C})$$

is complete and sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under the unrestricted multivariate normal family, as established previously.

Third, the conditional Stiefel representation shows that the conditional distribution of the sample given S is parameter-free. Therefore,

$$B_{\mathrm{RB}} = \mathbb{E}(B \mid S)$$

is an implementable function of the observed complete sufficient statistic.

Consequently, B_{RB} is an unbiased function of a complete sufficient statistic. By the Lehmann–Scheffé theorem, it is the unique uniformly minimum variance unbiased estimator of θ. □

Remark 7.2 (Scope of the UMVUE result). The UMVUE property established above is specific to the unrestricted multivariate normal family with unknown mean vector and unknown positive-definite covariance matrix.

The result does not establish that B_{RB} is a UMVUE over arbitrary nonparametric distributions. The completeness and sufficiency arguments rely essentially on the specified Gaussian model.

8 Variance Decomposition and Efficiency

The Rao–Blackwell estimator improves upon the original unbiased estimator in the sense of variance. This improvement can be quantified through the conditional variance decomposition.

Theorem 8.1 (Rao–Blackwell variance identity)**.** *Under the assumptions of the preceding theorem,*

$$\boxed{\mathrm{Var}(B) = \mathrm{Var}(B_{\mathrm{RB}}) + \mathbb{E}\{\mathrm{Var}(B \mid S)\}.}$$

Consequently,

$$\boxed{\mathrm{Var}(B_{\mathrm{RB}}) \leq \mathrm{Var}(B).}$$

Proof. Applying the law of total variance with respect to the complete sufficient statistic S, we have

$$\mathrm{Var}(B) = \mathrm{Var}\{\mathbb{E}(B \mid S)\} + \mathbb{E}\{\mathrm{Var}(B \mid S)\}.$$

Since

$$B_{\mathrm{RB}} = \mathbb{E}(B \mid S),$$

the identity becomes

$$\mathrm{Var}(B) = \mathrm{Var}(B_{\mathrm{RB}}) + \mathbb{E}\{\mathrm{Var}(B \mid S)\}.$$

The nonnegativity of conditional variance gives

$$\mathrm{Var}(B_{\mathrm{RB}}) \leq \mathrm{Var}(B).$$

□

The variance reduction achieved through Rao–Blackwellization is therefore

$$\boxed{\mathrm{Var}(B) - \mathrm{Var}(B_{\mathrm{RB}}) = \mathbb{E}\{\mathrm{Var}(B \mid S)\}.}$$

Whenever $\mathrm{Var}(B) > 0$, define the relative efficiency of the Rao–Blackwell estimator with respect to the original U-statistic by

$$\boxed{\mathrm{RE} = \frac{\mathrm{Var}(B)}{\mathrm{Var}(B_{\mathrm{RB}})},}$$

provided that $\mathrm{Var}(B_{\mathrm{RB}}) > 0$.

It follows that

$$\mathrm{RE} \geq 1.$$

Similarly, define the relative variance reduction by

$$\boxed{\mathrm{RVR} = \frac{\mathrm{Var}(B) - \mathrm{Var}(B_{\mathrm{RB}})}{\mathrm{Var}(B)}.}$$

Using the variance decomposition, this can equivalently be expressed as

$$\boxed{\mathrm{RVR} = \frac{\mathbb{E}\{\mathrm{Var}(B \mid S)\}}{\mathrm{Var}(B)}.}$$

Therefore,

$$0 \leq \mathrm{RVR} \leq 1.$$

These quantities provide a theoretical basis for comparing the efficiency of the original U-statistic and its Rao–Blackwellized version.

9 Relationship with the Nonparametric Hotelling Statistic

The preceding results establish the variance reduction obtained by conditioning on the complete sufficient statistic. We now investigate the relationship between this estimator and the conditional estimator based on the nonparametric Hotelling-type statistic.

Recall that

$$A = T_{\mathrm{NP}}^2, \qquad B_A = \mathbb{E}(B \mid A),$$

whereas

$$B_{\mathrm{RB}} = \mathbb{E}(B \mid S), \qquad S = (\overline{\mathbf{Z}}, \mathbf{C}).$$

Suppose that the specified nonparametric Hotelling statistic is constructed entirely from the complete sufficient statistic. In other words, assume that there exists a measurable function ϕ such that

$$\boxed{A = \phi(S).}$$

Then

$$\sigma(A) \subseteq \sigma(S).$$

Under this condition, conditioning on S uses at least as much sample information as conditioning on A alone.

Theorem 9.1 (Variance ordering of conditional estimators)**.** *Suppose that B is square-integrable and that A is a measurable function of the complete sufficient statistic S. Then*

$$\boxed{\mathrm{Var}(B_{\mathrm{RB}}) \leq \mathrm{Var}(B_A) \leq \mathrm{Var}(B).}$$

Proof. Since

$$\sigma(A) \subseteq \sigma(S),$$

the tower property of conditional expectation implies that

$$\begin{aligned} B_A &= \mathbb{E}(B \mid A) \\ &= \mathbb{E}\{\mathbb{E}(B \mid S) \mid A\} \\ &= \mathbb{E}(B_{\mathrm{RB}} \mid A). \end{aligned}$$

Applying the law of total variance to B_{RB} conditional on A, we obtain

$$\mathrm{Var}(B_{\mathrm{RB}}) = \mathrm{Var}\{\mathbb{E}(B_{\mathrm{RB}} \mid A)\} + \mathbb{E}\{\mathrm{Var}(B_{\mathrm{RB}} \mid A)\}.$$

Since

$$\mathbb{E}(B_{\mathrm{RB}} \mid A) = B_A,$$

it follows that

$$\mathrm{Var}(B_{\mathrm{RB}}) = \mathrm{Var}(B_A) + \mathbb{E}\{\mathrm{Var}(B_{\mathrm{RB}} \mid A)\}.$$

Therefore,

$$\mathrm{Var}(B_{\mathrm{RB}}) \leq \mathrm{Var}(B_A).$$

Similarly, applying the law of total variance to B conditional on A gives

$$\mathrm{Var}(B) = \mathrm{Var}(B_A) + \mathbb{E}\{\mathrm{Var}(B \mid A)\}.$$

Hence,

$$\mathrm{Var}(B_A) \leq \mathrm{Var}(B).$$

Combining the two inequalities yields

$$\boxed{\mathrm{Var}(B_{\mathrm{RB}}) \leq \mathrm{Var}(B_A) \leq \mathrm{Var}(B).}$$

□

Remark 9.2. The condition $A = \phi(S)$ is essential for the above variance ordering. If the nonparametric Hotelling statistic is not measurable with respect to the complete sufficient

statistic, the two conditioning σ-fields need not be nested, and the stated ordering between $\mathrm{Var}(B_A)$ and $\mathrm{Var}(B_{\mathrm{RB}})$ does not follow automatically.

Furthermore, although the conditional expectation B_A is unbiased, it may depend on unknown parameters when the conditional distribution of B given A is parameter-dependent. In contrast, the conditional distribution given the complete sufficient statistic is parameter-free under the assumed multivariate normal model.

10 Hoeffding Variance Decomposition

The original estimator

$$B = \widehat{\tau}^*_{p,q}$$

is a symmetric U-statistic of order four. Its variance can therefore be expressed through the Hoeffding decomposition [2].

This decomposition separates the sampling variability into components associated with projections of different orders. It also provides a useful theoretical framework for investigating the asymptotic behavior of the original U-statistic.

Let

$$h_0 = \theta$$

and define the first four Hoeffding projections recursively. In particular, for $c = 1, 2, 3, 4$, let

$$h_c(\mathbf{z}_1, \ldots, \mathbf{z}_c)$$

denote the canonical projection of order c. These projections are degenerate in the sense that, for $c \geq 1$,

$$\mathbb{E}\left[h_c(\mathbf{Z}_1, \ldots, \mathbf{Z}_c) \mid \mathbf{Z}_1, \ldots, \mathbf{Z}_{c-1}\right] = 0$$

almost surely, with the corresponding lower-order degeneracy conditions.

Define

$$\boxed{\sigma_c^2 = \mathbb{E}\left[h_c^2(\mathbf{Z}_1, \ldots, \mathbf{Z}_c)\right], \qquad c = 1, 2, 3, 4.}$$

Theorem 10.1 (Variance of the fourth-order U-statistic)**.** *Let B be a square-integrable symmetric U-statistic of order four based on iid observations. Then*

$$\boxed{\mathrm{Var}(B) = \sum_{c=1}^{4} \frac{\binom{4}{c}^2}{\binom{n}{c}} \sigma_c^2.}$$

Equivalently,

$$\boxed{\begin{aligned}\mathrm{Var}(B) = \frac{16}{n}\sigma_1^2 + \frac{36}{\binom{n}{2}}\sigma_2^2 \\ + \frac{16}{\binom{n}{3}}\sigma_3^2 + \frac{1}{\binom{n}{4}}\sigma_4^2.\end{aligned}}$$

Proof. The Hoeffding decomposition of a symmetric U-statistic of order four is

$$B - \theta = \sum_{c=1}^{4} \frac{\binom{4}{c}}{\binom{n}{c}} \sum_{1 \leq i_1 < \cdots < i_c \leq n} h_c(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_c}).$$

The canonical Hoeffding projections of different orders are mutually orthogonal in the L^2 sense. Moreover, the degeneracy property ensures that distinct terms of the same order are

uncorrelated unless they involve exactly the same set of observations.

Consequently, the variance of the Hoeffding decomposition is the sum of the variances of its individual components.

For a fixed order c, there are $\binom{n}{c}$ terms in the inner summation, and each term has variance σ_c^2. Thus, the contribution of the cth projection to the variance is

$$\left(\frac{\binom{4}{c}}{\binom{n}{c}}\right)^2 \binom{n}{c} \sigma_c^2 = \frac{\binom{4}{c}^2}{\binom{n}{c}} \sigma_c^2.$$

Summing the contributions over $c = 1, 2, 3, 4$ gives

$$\mathrm{Var}(B) = \sum_{c=1}^{4} \frac{\binom{4}{c}^2}{\binom{n}{c}} \sigma_c^2.$$

Finally, using

$$\binom{4}{1} = 4, \qquad \binom{4}{2} = 6, \qquad \binom{4}{3} = 4, \qquad \binom{4}{4} = 1,$$

we obtain

$$\begin{aligned} \mathrm{Var}(B) &= \frac{4^2}{\binom{n}{1}} \sigma_1^2 + \frac{6^2}{\binom{n}{2}} \sigma_2^2 \\ &\quad + \frac{4^2}{\binom{n}{3}} \sigma_3^2 + \frac{1^2}{\binom{n}{4}} \sigma_4^2 \\ &= \frac{16}{n} \sigma_1^2 + \frac{36}{\binom{n}{2}} \sigma_2^2 \\ &\quad + \frac{16}{\binom{n}{3}} \sigma_3^2 + \frac{1}{\binom{n}{4}} \sigma_4^2. \end{aligned}$$

This proves the result. □

The Hoeffding decomposition describes the variance of the original fourth-order U-statistic, whereas the Rao–Blackwell variance identity describes the reduction achieved by conditioning on the complete sufficient statistic. Combining these two representations gives

$$\boxed{\begin{aligned} &\frac{16}{n} \sigma_1^2 + \frac{36}{\binom{n}{2}} \sigma_2^2 + \frac{16}{\binom{n}{3}} \sigma_3^2 + \frac{1}{\binom{n}{4}} \sigma_4^2 \\ &\qquad = \mathrm{Var}(B_{\mathrm{RB}}) + \mathbb{E}\{\mathrm{Var}(B \mid S)\}. \end{aligned}}$$

This identity provides a theoretical connection between the classical variance decomposition of U-statistics and the variance improvement obtained through Rao–Blackwellization.

11 Monte Carlo Approximation

11.1 Motivation and Computational Formulation

The Rao–Blackwell estimator obtained in the preceding section is expressed as an integral over the constrained Stiefel manifold:

$$B_{\mathrm{RB}} = \int_{\mathcal{V}_{d,n-1}} B(\mathbf{Q})\, d\nu_{d,n-1}(\mathbf{Q}),$$

where

$$B(\mathbf{Q}) = \widehat{\tau}^{*}_{p,q}\left(\mathbf{Z}_1(\mathbf{Q}), \ldots, \mathbf{Z}_n(\mathbf{Q})\right),$$

and

$$\mathbf{Z}_i(\mathbf{Q}) = \overline{\mathbf{Z}} + \mathbf{C}^{1/2}\mathbf{q}_i.$$

Here, $\mathbf{q}_i$ denotes the transpose of the ith row of $\mathbf{Q}$, and $\nu_{d,n-1}$ is the uniform invariant probability measure on

$$\mathcal{V}_{d,n-1} = \left\{\mathbf{Q} \in \mathbb{R}^{n\times d} : \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_d,\ \mathbf{1}_n^T\mathbf{Q} = \mathbf{0}^T\right\}.$$

Although this representation is analytically explicit, the evaluation of the integral may be computationally demanding. In particular, the dimension of the constrained Stiefel manifold increases with both the sample size and the dimension of the joint observation vector. Furthermore, the integrand involves a fourth-order U-statistic whose kernel contains integrations over the projection directions belonging to two unit spheres.

Consequently, a direct analytical evaluation of the conditional integral may not be feasible in general. We therefore consider a Monte Carlo approximation based on independent observations from the uniform invariant distribution on the constrained Stiefel manifold.

11.2 Monte Carlo Construction

Let

$$\mathbf{Q}^{(1)}, \mathbf{Q}^{(2)}, \ldots, \mathbf{Q}^{(M)} \overset{\text{iid}}{\sim} \nu_{d,n-1}$$

be independent random matrices generated from the uniform invariant probability measure on $\mathcal{V}_{d,n-1}$.

For each $m = 1, \ldots, M$, construct the corresponding sample configuration

$$\mathbf{Z}_i^{(m)} = \overline{\mathbf{Z}} + \mathbf{C}^{1/2}\mathbf{q}_i^{(m)}, \qquad i = 1, \ldots, n,$$

where $\mathbf{q}_i^{(m)}$ is the transpose of the ith row of $\mathbf{Q}^{(m)}$.

The resulting fourth-order U-statistic is

$$B(\mathbf{Q}^{(m)}) = \binom{n}{4}^{-1} \sum_{1\le i<j<k<l\le n} h\left(\mathbf{Z}_i^{(m)}, \mathbf{Z}_j^{(m)}, \mathbf{Z}_k^{(m)}, \mathbf{Z}_l^{(m)}\right).$$

The Monte Carlo approximation to the exact Rao–Blackwell estimator is defined by

$$\boxed{\widehat{B}_{\mathrm{RB},M} = \frac{1}{M}\sum_{m=1}^{M} B(\mathbf{Q}^{(m)}).} \tag{3}$$

Equivalently,

$$\boxed{\widehat{B}_{\mathrm{RB},M} = \frac{1}{M\binom{n}{4}} \sum_{m=1}^{M} \sum_{1 \le i < j < k < l \le n} h\left(\mathbf{Z}_i^{(m)}, \mathbf{Z}_j^{(m)}, \mathbf{Z}_k^{(m)}, \mathbf{Z}_l^{(m)}\right).} \tag{4}$$

Thus, the proposed numerical procedure consists of generating M sample configurations that share the same observed sample mean and scatter matrix, but have independently generated orientations on the constrained Stiefel manifold. The corresponding U-statistic values are then averaged to approximate their conditional expectation.

11.3 Conditional Unbiasedness

The first important property of the Monte Carlo approximation is its conditional unbiasedness for the exact Rao–Blackwell estimator.

Theorem 11.1 (Conditional unbiasedness of the Monte Carlo approximation)**.** *Suppose that*

$$\mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(M)} \overset{\text{iid}}{\sim} \nu_{d,n-1}$$

are generated independently from the conditional Stiefel distribution given

$$S = (\overline{\mathbf{Z}}, \mathbf{C}).$$

Then

$$\boxed{\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) = B_{\mathrm{RB}}.}$$

Proof. By definition,

$$\widehat{B}_{\mathrm{RB},M} = \frac{1}{M} \sum_{m=1}^{M} B(\mathbf{Q}^{(m)}).$$

Taking conditional expectations given S,

$$\begin{aligned} \mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) &= \mathbb{E}\left[\frac{1}{M} \sum_{m=1}^{M} B(\mathbf{Q}^{(m)}) \mid S\right] \\ &= \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left[B(\mathbf{Q}^{(m)}) \mid S\right]. \end{aligned}$$

Since each $\mathbf{Q}^{(m)}$ has the uniform conditional distribution on $\mathcal{V}_{d,n-1}$, the conditional expectation of every summand is identical. In particular,

$$\begin{aligned} \mathbb{E}\left[B(\mathbf{Q}^{(m)}) \mid S\right] &= \int_{\mathcal{V}_{d,n-1}} B(\mathbf{Q})\, d\nu_{d,n-1}(\mathbf{Q}) \\ &= B_{\mathrm{RB}}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) &= \frac{1}{M} \sum_{m=1}^{M} B_{\mathrm{RB}} \\ &= B_{\mathrm{RB}}. \end{aligned}$$

This establishes the conditional unbiasedness. □

11.4 Conditional Monte Carlo Variance

The numerical approximation introduces an additional source of variability arising from the random generation of the Stiefel matrices. This variability can be quantified exactly.

Theorem 11.2 (Conditional variance of the Monte Carlo approximation)**.** *Under the assumptions of the preceding theorem, if B is square-integrable, then*

$$\boxed{\operatorname{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) = \frac{1}{M} \operatorname{Var}(B \mid S).}$$

Proof. Using the definition of the Monte Carlo estimator,

$$\operatorname{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) = \operatorname{Var}\left(\frac{1}{M} \sum_{m=1}^{M} B(\mathbf{Q}^{(m)}) \mid S \right).$$

Since the matrices $\mathbf{Q}^{(1)}, \ldots, \mathbf{Q}^{(M)}$ are conditionally independent given S, the corresponding random variables $B(\mathbf{Q}^{(1)}), \ldots, B(\mathbf{Q}^{(M)})$ are also conditionally independent.

Consequently, all conditional covariance terms between distinct summands vanish. Hence,

$$\begin{aligned} \operatorname{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) &= \frac{1}{M^2} \sum_{m=1}^{M} \operatorname{Var}(B(\mathbf{Q}^{(m)}) \mid S) \\ &= \frac{1}{M^2} \sum_{m=1}^{M} \operatorname{Var}(B \mid S). \end{aligned}$$

Since there are M identical terms,

$$\begin{aligned} \operatorname{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) &= \frac{M}{M^2} \operatorname{Var}(B \mid S) \\ &= \frac{1}{M} \operatorname{Var}(B \mid S). \end{aligned}$$

The result follows. □

The theorem shows that the additional conditional simulation variance decreases at the rate M^{-1}. In particular, increasing the number of generated Stiefel matrices reduces the numerical uncertainty associated with approximating the exact conditional expectation.

11.5 Mean-Square Error of the Monte Carlo Approximation

Since the Monte Carlo approximation is conditionally unbiased for B_{RB}, its conditional mean-square error is equal to its conditional variance.

Corollary 11.3 (Conditional mean-square error)**.** *Under the preceding assumptions,*

$$\boxed{\begin{aligned} &\mathbb{E}\left[(\widehat{B}_{\mathrm{RB},M} - B_{\mathrm{RB}})^2 \mid S \right] \\ &\qquad = \frac{1}{M} \operatorname{Var}(B \mid S). \end{aligned}}$$

Proof. By the conditional mean-square error decomposition,

$$\begin{aligned}&\mathbb{E}\left[(\widehat{B}_{\mathrm{RB},M} - B_{\mathrm{RB}})^2 \mid S\right] \\ &= \mathrm{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) + \left[\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) - B_{\mathrm{RB}}\right]^2.\end{aligned}$$

Conditional unbiasedness gives

$$\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) - B_{\mathrm{RB}} = 0.$$

Therefore,

$$\begin{aligned}&\mathbb{E}\left[(\widehat{B}_{\mathrm{RB},M} - B_{\mathrm{RB}})^2 \mid S\right] \\ &\quad = \mathrm{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) = \frac{1}{M}\,\mathrm{Var}(B \mid S).\end{aligned}$$

□

Taking expectations with respect to the distribution of S gives

$$\boxed{\mathbb{E}\left[(\widehat{B}_{\mathrm{RB},M} - B_{\mathrm{RB}})^2\right] = \frac{1}{M}\mathbb{E}\{\mathrm{Var}(B \mid S)\}.} \tag{5}$$

Thus, the unconditional mean-square simulation error also decreases at the rate M^{-1}.

11.6 Unconditional Unbiasedness

The Monte Carlo approximation is also unbiased for the original population dependence parameter when both the sampling mechanism and the Monte Carlo randomization are taken into account.

Theorem 11.4 (Unconditional unbiasedness)**.** *Suppose that B is unbiased for θ and that the Stiefel matrices are generated according to the conditional distribution specified above. Then*

$$\boxed{\mathbb{E}(\widehat{B}_{\mathrm{RB},M}) = \theta.}$$

Proof. By the law of iterated expectations,

$$\begin{aligned}\mathbb{E}(\widehat{B}_{\mathrm{RB},M}) &= \mathbb{E}\left[\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S)\right] \\ &= \mathbb{E}(B_{\mathrm{RB}}).\end{aligned}$$

Again, applying the law of iterated expectations,

$$\begin{aligned}\mathbb{E}(B_{\mathrm{RB}}) &= \mathbb{E}\{\mathbb{E}(B \mid S)\} \\ &= \mathbb{E}(B).\end{aligned}$$

Since B is unbiased for θ,

$$\mathbb{E}(B) = \theta.$$

Therefore,

$$\boxed{\mathbb{E}(\widehat{B}_{\mathrm{RB},M}) = \theta.}$$

□

This result distinguishes the finite-simulation estimator from the exact Rao–Blackwell estimator: although the former contains additional simulation variability, it remains unbiased for the target parameter when the Monte Carlo randomization is included in the expectation.

11.7 Unconditional Variance Decomposition

The variance of the Monte Carlo approximation contains two components: the sampling variability of the exact Rao–Blackwell estimator and the additional variability introduced by Monte Carlo integration.

Theorem 11.5 (Unconditional variance of the Monte Carlo approximation)**.** *Suppose that the Stiefel matrices are conditionally independent given S. Then*

$$\boxed{\operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) = \operatorname{Var}(B_{\mathrm{RB}}) + \frac{1}{M}\mathbb{E}\{\operatorname{Var}(B \mid S)\}.}$$

Equivalently,

$$\boxed{\operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) = \operatorname{Var}(B_{\mathrm{RB}}) + \frac{\operatorname{Var}(B) - \operatorname{Var}(B_{\mathrm{RB}})}{M}.}$$

Proof. Applying the law of total variance with respect to S,

$$\begin{aligned}\operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) = {} & \operatorname{Var}\left\{\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S)\right\} \\ & + \mathbb{E}\left\{\operatorname{Var}(\widehat{B}_{\mathrm{RB},M} \mid S)\right\}.\end{aligned}$$

By conditional unbiasedness,

$$\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) = B_{\mathrm{RB}}.$$

Furthermore, by the conditional variance result,

$$\operatorname{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) = \frac{1}{M}\operatorname{Var}(B \mid S).$$

Substituting these expressions gives

$$\operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) = \operatorname{Var}(B_{\mathrm{RB}}) + \frac{1}{M}\mathbb{E}\{\operatorname{Var}(B \mid S)\}.$$

This establishes the first expression.
Now, from the Rao–Blackwell variance identity,

$$\operatorname{Var}(B) = \operatorname{Var}(B_{\mathrm{RB}}) + \mathbb{E}\{\operatorname{Var}(B \mid S)\}.$$

Therefore,

$$\mathbb{E}\{\operatorname{Var}(B \mid S)\} = \operatorname{Var}(B) - \operatorname{Var}(B_{\mathrm{RB}}).$$

Substitution yields

$$\operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) = \operatorname{Var}(B_{\mathrm{RB}}) + \frac{\operatorname{Var}(B) - \operatorname{Var}(B_{\mathrm{RB}})}{M}.$$

This completes the proof. □

Remark 11.6 (Interpretation of the variance decomposition). The preceding identity shows that

$$\operatorname{Var}(B_{\mathrm{RB}}) \le \operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) \le \operatorname{Var}(B), \qquad M \ge 1.$$

For $M = 1$, the Monte Carlo estimator has the same unconditional variance as the original U-statistic. As M increases, its variance decreases monotonically towards the variance of the exact Rao–Blackwell estimator.

In particular,

$$\lim_{M\to\infty} \operatorname{Var}(\widehat{B}_{\mathrm{RB},M}) = \operatorname{Var}(B_{\mathrm{RB}}).$$

Thus, the Monte Carlo procedure gradually recovers the variance improvement of the exact Rao–Blackwell estimator.

11.8 Consistency of the Monte Carlo Approximation

The Monte Carlo approximation converges to the exact conditional expectation as the number of generated Stiefel matrices tends to infinity.

Theorem 11.7 (Conditional strong consistency)**.** *Suppose that*

$$\mathbb{E}(|B| \mid S) < \infty$$

almost surely and that the Stiefel matrices are conditionally iid given S. Then

$$\boxed{\widehat{B}_{\mathrm{RB},M} \longrightarrow B_{\mathrm{RB}} \qquad \textit{almost surely as } M \to \infty.}$$

Proof. For a fixed observed sufficient statistic S, the random variables

$$B(\mathbf{Q}^{(1)}), \dots, B(\mathbf{Q}^{(M)})$$

are conditionally iid and integrable.

Their common conditional expectation is

$$\mathbb{E}\{B(\mathbf{Q}^{(m)}) \mid S\} = B_{\mathrm{RB}}.$$

Therefore, the conditional strong law of large numbers gives

$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M} B(\mathbf{Q}^{(m)}) &\longrightarrow \mathbb{E}\{B(\mathbf{Q}^{(1)}) \mid S\} \\ &= B_{\mathrm{RB}}\end{aligned}$$

almost surely under the conditional simulation distribution.

Since the left-hand side is precisely $\widehat{B}_{\mathrm{RB},M}$, we obtain

$$\boxed{\widehat{B}_{\mathrm{RB},M} \longrightarrow B_{\mathrm{RB}} \qquad \text{almost surely.}}$$

□

In the present setting, the kernel is bounded. Consequently, the U-statistic B is bounded, which guarantees the required conditional integrability.

11.9 Monte Carlo Standard Error

For practical implementation, the conditional simulation variance can be estimated using the sample variance of the simulated U-statistic values.

Define

$$\overline{B}_M = \frac{1}{M}\sum_{m=1}^{M} B(\mathbf{Q}^{(m)})$$

and

$$\boxed{s_M^2 = \frac{1}{M-1}\sum_{m=1}^{M}\left\{B(\mathbf{Q}^{(m)}) - \overline{B}_M\right\}^2.}$$

Then the conditional Monte Carlo variance is estimated by

$$\boxed{\widehat{\mathrm{Var}}_{\mathrm{MC}} = \frac{s_M^2}{M}.}$$

Accordingly, the estimated conditional Monte Carlo standard error is

$$\boxed{\widehat{\mathrm{MCSE}} = \frac{s_M}{\sqrt{M}}.}$$

This quantity measures the numerical uncertainty in approximating the exact Rao–Blackwell conditional expectation for the observed value of S.

It is important to distinguish this simulation standard error from the sampling standard error of the estimator itself. The former quantifies the error arising from a finite number of Stiefel simulations, whereas the latter concerns the variability of the estimator over repeated samples from the population distribution.

11.10 Generation of Uniform Stiefel Matrices

The Monte Carlo approximation requires independent random matrices from the uniform invariant probability measure on

$$\mathcal{V}_{d,n-1} = \left\{\mathbf{Q} \in \mathbb{R}^{n\times d} : \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_d,\ \mathbf{1}_n^T\mathbf{Q} = \mathbf{0}^T\right\}.$$

A convenient construction is based on Gaussian random matrices and an orthonormalization procedure.

Define the centering projection matrix

$$\boxed{\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T.}$$

This matrix satisfies

$$\mathbf{H}^T = \mathbf{H}, \qquad \mathbf{H}^2 = \mathbf{H}, \qquad \mathbf{H}\mathbf{1}_n = \mathbf{0}.$$

It is the orthogonal projection onto the $(n-1)$-dimensional subspace

$$\mathcal{L} = \left\{\mathbf{u} \in \mathbb{R}^n : \mathbf{1}_n^T\mathbf{u} = 0\right\}.$$

The following procedure generates a uniform random element of the constrained Stiefel manifold.

1. Generate an $n \times d$ random matrix

$$\mathbf{G} = (G_{ij}),$$

where the entries are mutually independent standard normal random variables:

$$G_{ij} \overset{\text{iid}}{\sim} N(0,1).$$

2. Center the Gaussian matrix by calculating

$$\mathbf{W} = \mathbf{H}\mathbf{G}.$$

The columns of $\mathbf{W}$ belong to the subspace $\mathcal{L}$.

3. Compute a thin QR decomposition:

$$\mathbf{W} = \mathbf{Q}\mathbf{R},$$

where $\mathbf{Q}$ is an $n \times d$ matrix with orthonormal columns and $\mathbf{R}$ is a $d \times d$ upper triangular matrix.

4. The resulting matrix $\mathbf{Q}$ satisfies

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_d, \qquad \mathbf{1}_n^T\mathbf{Q} = \mathbf{0}^T.$$

Consequently,

$$\mathbf{Q} \in \mathcal{V}_{d,n-1}.$$

Under the Gaussian construction, the resulting orientation is uniformly distributed on the constrained Stiefel manifold.

11.11 Justification of the Uniform Distribution

Lemma 11.8 (Uniformity of the Gaussian–QR construction)**.** *Let* $\mathbf{G}$ *be an* $n \times d$ *matrix with independent standard normal entries, and let*

$$\mathbf{W} = \mathbf{H}\mathbf{G}.$$

Suppose that $n - 1 \geq d$*. If*

$$\mathbf{W} = \mathbf{Q}\mathbf{R}$$

is a thin QR decomposition with a fixed consistent sign convention, then $\mathbf{Q}$ *has the uniform invariant distribution on* $\mathcal{V}_{d,n-1}$*.*

Proof. Let $\mathbf{U}$ be an $n \times (n-1)$ matrix whose columns form an orthonormal basis of $\mathcal{L}$. Thus,

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_{n-1}, \qquad \mathbf{U}^T\mathbf{1}_n = \mathbf{0}, \qquad \mathbf{U}\mathbf{U}^T = \mathbf{H}.$$

Since

$$\mathbf{H}\mathbf{G} = \mathbf{U}\mathbf{U}^T\mathbf{G},$$

we can write

$$\mathbf{W} = \mathbf{U}\mathbf{A}, \qquad \mathbf{A} = \mathbf{U}^T\mathbf{G}.$$

Because $\mathbf{G}$ has independent standard normal entries and $\mathbf{U}$ has orthonormal columns, the entries of

$$\mathbf{A} \in \mathbb{R}^{(n-1)\times d}$$

are also independent standard normal random variables.

Consider the thin QR decomposition

$$\mathbf{A} = \mathbf{Q}_0\mathbf{R}.$$

Then

$$\mathbf{W} = \mathbf{U}\mathbf{Q}_0\mathbf{R}.$$

Therefore, the orthonormal factor in the QR decomposition of $\mathbf{W}$ is

$$\mathbf{Q} = \mathbf{U}\mathbf{Q}_0.$$

The standard Gaussian matrix $\mathbf{A}$ is invariant in distribution under left multiplication by any orthogonal matrix $\mathbf{O} \in \mathbb{R}^{(n-1)\times(n-1)}$:

$$\mathbf{O}\mathbf{A} \overset{d}{=} \mathbf{A}.$$

Consequently, the orthonormal factor $\mathbf{Q}_0$ has the uniform invariant distribution on the ordinary Stiefel manifold

$$\left\{\mathbf{Q}_0 \in \mathbb{R}^{(n-1)\times d} : \mathbf{Q}_0^T\mathbf{Q}_0 = \mathbf{I}_d\right\}.$$

The mapping

$$\mathbf{Q}_0 \longmapsto \mathbf{U}\mathbf{Q}_0$$

is an isometry from this Stiefel manifold onto $\mathcal{V}_{d,n-1}$. Hence,

$$\mathbf{Q} = \mathbf{U}\mathbf{Q}_0$$

has the uniform invariant distribution on $\mathcal{V}_{d,n-1}$.

This establishes the result. □

Remark 11.9 (Rank condition). The Gaussian–QR construction requires

$$n - 1 \geq d.$$

The standing assumption $n > d$ therefore ensures that the centered Gaussian matrix has full column rank with probability one.

In numerical implementation, a rank-deficient QR factorization should be treated as a numerical failure, although it occurs with probability zero under the ideal continuous Gaussian model.

11.12 Alternative Polar Decomposition

An alternative way of obtaining a uniformly distributed Stiefel matrix is through the polar decomposition.

Let

$$\mathbf{W} = \mathbf{H}\mathbf{G}$$

and suppose that $\mathbf{W}$ has full column rank. Define

$$\boxed{\mathbf{Q} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1/2}.}$$

Then

$$\begin{aligned}\mathbf{Q}^T\mathbf{Q} &= (\mathbf{W}^T\mathbf{W})^{-1/2}\mathbf{W}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1/2} \\ &= \mathbf{I}_d.\end{aligned}$$

Moreover,

$$\mathbf{1}_n^T\mathbf{Q} = \mathbf{1}_n^T\mathbf{H}\mathbf{G}(\mathbf{W}^T\mathbf{W})^{-1/2} = \mathbf{0}^T.$$

Thus,

$$\mathbf{Q} \in \mathcal{V}_{d,n-1}.$$

The polar construction produces the uniform invariant Stiefel distribution because the Gaussian distribution is orthogonally invariant. It is particularly useful when a symmetric orthonormalization is preferred over a triangular QR factorization.

11.13 Practical Monte Carlo Algorithm

The complete computational procedure can be summarized as follows.

1. Calculate the observed sample mean $\overline{\mathbf{Z}}$ and sample scatter matrix $\mathbf{C}$.
2. Obtain the symmetric positive-definite square root $\mathbf{C}^{1/2}$.
3. Construct the centering matrix

$$\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T.$$

4. For $m = 1, \ldots, M$:
   (a) Generate an $n \times d$ standard Gaussian matrix $\mathbf{G}^{(m)}$.
   (b) Calculate

$$\mathbf{W}^{(m)} = \mathbf{H}\mathbf{G}^{(m)}.$$

   (c) Compute the thin QR decomposition

$$\mathbf{W}^{(m)} = \mathbf{Q}^{(m)}\mathbf{R}^{(m)}.$$

   (d) Construct the conditional sample

$$\mathbf{Z}_i^{(m)} = \overline{\mathbf{Z}} + \mathbf{C}^{1/2}\mathbf{q}_i^{(m)}.$$

   (e) Evaluate the fourth-order U-statistic

$$B(\mathbf{Q}^{(m)}) = \binom{n}{4}^{-1} \sum_{i<j<k<l} h\left(\mathbf{Z}_i^{(m)}, \mathbf{Z}_j^{(m)}, \mathbf{Z}_k^{(m)}, \mathbf{Z}_l^{(m)}\right).$$

5. Calculate the Monte Carlo approximation

$$\widehat{B}_{\mathrm{RB},M} = \frac{1}{M}\sum_{m=1}^{M} B(\mathbf{Q}^{(m)}).$$

6. Estimate the Monte Carlo standard error by

$$\widehat{\mathrm{MCSE}} = \frac{s_M}{\sqrt{M}}.$$

Remark 11.10 (Computational complexity). For each generated Stiefel matrix, evaluating the fourth-order U-statistic requires $\binom{n}{4}$ kernel evaluations. Thus, a direct implementation has computational cost proportional to

$$O\left(M\binom{n}{4}\right)$$

kernel evaluations, apart from the additional cost associated with the projection integrations.

For large sample sizes, this cost may become substantial. Possible computational improvements include parallel generation of the Stiefel matrices, vectorization of the kernel evaluations, reusing projection directions, and numerical approximation of the sphere integrals.

Such computational approximations should be distinguished from the exact mathematical definition of the Rao–Blackwell estimator.

11.14 Results

The numerical performance of the proposed conditional estimation procedure was investigated through a Monte Carlo simulation study. The experiment was designed to examine the computational validity of the constrained Stiefel representation and to compare the sampling performance of the original fourth-order U-statistic with its Monte Carlo Rao–Blackwell approximation.

Throughout the experiment, the joint observation vector was generated from the multivariate normal distribution

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} \sim N_4(\mathbf{0}, \mathbf{\Sigma}), \qquad i = 1, \dots, n,$$

where $\mathbf{X}_i \in \mathbb{R}^2$ and $\mathbf{Y}_i \in \mathbb{R}^2$. The marginal covariance matrices were taken to be identity matrices, while the cross-covariance matrix was specified as

$$\mathbf{\Sigma}_{XY} = \begin{pmatrix} \rho & 0 \\ 0 & 0 \end{pmatrix}.$$

Consequently, the joint covariance matrix was

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & \rho & 0 \\ 0 & 1 & 0 & 0 \\ \rho & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

with $\rho = 0.50$ in the principal simulation experiment. This choice produces a positive-definite covariance matrix and introduces dependence between the first coordinates of $\mathbf{X}_i$ and $\mathbf{Y}_i$.

The sample size was fixed at $n = 12$, and the projection-based kernel was approximated using $L = 100$ independently generated pairs of projection directions. For the conditional Stiefel integration, $M = 50$ independent constrained Stiefel matrices were generated for each observed dataset. The repeated-sampling experiment consisted of $R = 200$ independent datasets. The random-number generator was initialized using the seed 20261105 to ensure reproducibility.

For each generated dataset, the original estimator was calculated as

$$B = \frac{1}{\binom{n}{4}} \sum_{1 \le i_1 < i_2 < i_3 < i_4 \le n} \widehat{h}_L(\mathbf{Z}_{i_1}, \mathbf{Z}_{i_2}, \mathbf{Z}_{i_3}, \mathbf{Z}_{i_4}),$$

where $\widehat{h}_L$ denotes the Monte Carlo approximation to the projection-integrated kernel. The cor-

responding Rao–Blackwell approximation was obtained by averaging the values of the original statistic over M independently generated conditional Stiefel configurations:

$$\widehat{B}_{\mathrm{RB},M} = \frac{1}{M}\sum_{m=1}^{M} B(\mathbf{Q}^{(m)}).$$

11.14.1 Verification of the Stiefel Representation

The first numerical investigation concerned the validity of the generated constrained Stiefel matrices. Each matrix $\mathbf{Q}$ was required to satisfy

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_d, \qquad \mathbf{1}_n^T\mathbf{Q} = \mathbf{0}^T.$$

The numerical implementation was based on the polar orthonormalization of a centered Gaussian matrix. The resulting matrix was checked against both defining constraints.

For the simulated configuration with $n = 12$ and $d = 4$, the maximum absolute errors were

$$\max_{i,j} \left|(\mathbf{Q}^T\mathbf{Q} - \mathbf{I}_4)_{ij}\right| = 1.110223 \times 10^{-15}$$

and

$$\max_{j} \left|\sum_{i=1}^{12} Q_{ij}\right| = 1.533062 \times 10^{-16}.$$

Both errors are of the order of machine precision. Thus, the generated matrices satisfy the orthogonality and centering restrictions numerically.

The reconstruction procedure was subsequently examined by generating a conditional sample

$$\mathbf{Z}^* = \mathbf{1}_n\overline{\mathbf{Z}}^T + \mathbf{Q}\mathbf{C}^{1/2},$$

where $\overline{\mathbf{Z}}$ and $\mathbf{C}$ denote the observed sample mean and sample scatter matrix, respectively. The reconstructed sample was required to preserve these two quantities. The numerical differences between the original and reconstructed sufficient statistics were

$$\max_{j} \left|\overline{Z}_j - \overline{Z}_j^*\right| = 5.551115 \times 10^{-17}$$

and

$$\max_{i,j} \left|(C - C^*)_{ij}\right| = 2.309264 \times 10^{-14}.$$

These results confirm that the proposed reconstruction preserves the observed sample mean and scatter matrix up to numerical rounding errors. Therefore, the computational implementation is consistent with the conditional Stiefel representation developed theoretically.

11.14.2 Comparison of the Original and Rao–Blackwell Estimators

For an illustrative observed dataset, the original projection-based U-statistic was

$$B = 0.06818182,$$

whereas the Monte Carlo Rao–Blackwell estimator was

$$\widehat{B}_{\mathrm{RB},50} = 0.04292970.$$

The estimated conditional Monte Carlo standard error was

$$\widehat{\mathrm{MCSE}} = 0.003152813,$$

and the estimated conditional simulation variance of the Monte Carlo average was

$$\widehat{\mathrm{Var}}_{\mathrm{MC}} = 9.940231 \times 10^{-6}.$$

The difference between the two estimates for this particular dataset is not unexpected. The original statistic represents one observed sample configuration, whereas the conditional estimator averages the statistic over multiple sample configurations having the same sample mean and scatter matrix. Consequently, the Rao–Blackwell approximation reflects the conditional average over the corresponding Stiefel manifold rather than the value associated with the original configuration alone.

An independent Monte Carlo calculation was also performed to obtain a numerical reference for the population dependence parameter. The approximate reference value was

$$\widehat{\theta}_{\mathrm{ref}} = 0.02045350,$$

with estimated Monte Carlo standard error

$$\widehat{\mathrm{SE}}(\widehat{\theta}_{\mathrm{ref}}) = 0.001752326.$$

This value was subsequently used as the reference parameter in the repeated-sampling comparison. It should be interpreted as a numerical approximation rather than an exact analytical evaluation of the population dependence parameter.

11.14.3 Repeated-Sampling Performance

The sampling performance of the two estimators was evaluated using $R = 200$ independent replications. For each replication, the original U-statistic and its Monte Carlo Rao–Blackwell approximation were calculated using the same projection-direction collection. The empirical bias, variance, root mean squared error (RMSE), and mean absolute error (MAE) were then computed relative to the numerical reference parameter.

The resulting performance measures are reported in Table 1.

Table 1: Repeated-sampling performance of the original and Monte Carlo Rao–Blackwell estimators.

Estimator	Mean	Bias	Variance	RMSE	MAE
Original U-statistic	0.02079071	0.00033721	0.00174610	0.04168315	0.03238881
Monte Carlo Rao–Blackwell	0.01988780	−0.00056570	0.00090798	0.03006261	0.02377579

The empirical mean of the original U-statistic was 0.02079071, while that of the Monte Carlo Rao–Blackwell estimator was 0.01988780. Their corresponding empirical biases were 0.00033721 and −0.00056570, respectively. Both estimators therefore exhibited relatively small empirical biases with respect to the numerical reference parameter.

A more pronounced difference was observed in their sampling variances. The variance of the original U-statistic was estimated as

$$\widehat{\mathrm{Var}}(B) = 0.0017461022,$$

whereas the corresponding variance of the Monte Carlo Rao–Blackwell estimator was

$$\widehat{\mathrm{Var}}(\widehat{B}_{\mathrm{RB},50}) = 0.0009079804.$$

The empirical relative efficiency, defined as the ratio of the variance of the original estimator to that of the Rao–Blackwell approximation, was therefore

$$\begin{aligned}\widehat{\mathrm{RE}} &= \frac{\widehat{\mathrm{Var}}(B)}{\widehat{\mathrm{Var}}(\widehat{B}_{\mathrm{RB},50})} \\ &= \frac{0.0017461022}{0.0009079804} \\ &\approx 1.9231.\end{aligned}$$

The corresponding percentage reduction in empirical variance was

$$\begin{aligned}\widehat{\mathrm{VR}} &= 100\left[1 - \frac{\widehat{\mathrm{Var}}(\widehat{B}_{\mathrm{RB},50})}{\widehat{\mathrm{Var}}(B)}\right] \\ &\approx 48.00\%.\end{aligned}$$

Thus, under the specified simulation configuration, the Monte Carlo Rao–Blackwell estimator exhibited approximately 48% lower empirical variance than the original U-statistic. Equivalently, the empirical variance of the original estimator was approximately 1.92 times that of the Rao–Blackwell approximation.

The RMSE also decreased from 0.04168315 for the original estimator to 0.03006261 for the conditional estimator. Similarly, the MAE decreased from 0.03238881 to 0.02377579. These results indicate that the observed reduction in sampling variance was accompanied by an improvement in the overall estimation accuracy in the present experiment.

It is important to emphasize that the reported relative efficiency and variance reduction are empirical quantities based on a finite number of repeated samples and a finite number of conditional Stiefel simulations. They therefore provide numerical evidence for the potential variance-reduction benefit of the proposed construction, rather than an exact numerical evaluation of the variance difference between the original estimator and its Rao–Blackwell counterpart.

11.14.4 Sensitivity to the Number of Projection Directions

The projection-integrated kernel was approximated using a finite number L of independently generated pairs of projection directions. Accordingly, the accuracy of the numerical approximation may depend on the choice of L. To investigate this dependence, the original U-statistic was evaluated for several values of L, namely $L \in \{25, 50, 100, 250, 500, 1000\}$. The resulting estimates for a fixed observed dataset are presented in Table 2.

Table 2: Sensitivity of the projection-based estimator to the number of projection-direction pairs.

Number of projection pairs (L)	Estimated statistic
25	0.08816162
50	0.10060606
100	0.07090909
250	0.06610909
500	0.07150707
1000	0.07072525

The numerical estimates exhibit noticeable variation for smaller values of L. In particular, the estimates corresponding to $L = 25$ and $L = 50$ differ appreciably from those obtained with larger projection collections. For $L = 250$, 500, and 1000, the estimates are comparatively closer, although some variation remains.

These observations suggest that the numerical approximation of the projection integral requires a sufficiently large number of projection directions. Nevertheless, the reported experiment should be regarded as a preliminary sensitivity investigation, because a new collection of random projection directions was generated for each value of L. Consequently, the differences between the estimates reflect both finite-simulation integration error and the randomness of the generated projection directions. A more controlled assessment would employ nested projection-direction collections and repeated independent direction sets.

11.14.5 Summary of the Numerical Findings

The simulation results provide numerical support for the computational feasibility of the constrained Stiefel representation. The orthogonality and centering restrictions were satisfied to machine precision, and the reconstructed samples preserved the observed sample mean and scatter matrix with negligible numerical error.

The repeated-sampling comparison further indicated that the Monte Carlo Rao–Blackwell approximation achieved lower empirical variance, RMSE, and MAE than the original fourth-order U-statistic under the considered simulation configuration. The observed empirical variance reduction was approximately 48%.

Finally, the sensitivity analysis demonstrated that the numerical approximation to the projection integral may be affected by the number of projection-direction pairs. Therefore, the selection of both the projection-simulation size and the conditional Stiefel-simulation size is important for obtaining reliable numerical results.

11.15 Finite-Simulation Interpretation

For finite M, the quantity $\widehat{B}_{\mathrm{RB},M}$ is a Monte Carlo approximation to the exact Rao–Blackwell estimator

$$B_{\mathrm{RB}} = \mathbb{E}(B \mid S),$$

rather than the exact conditional expectation itself. In particular, conditional on the observed sufficient statistic S and on the projection-direction collection used to define the numerical kernel,

$$\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) = B_{\mathrm{RB}},$$

where the expectation is taken with respect to the independent Stiefel simulations. However, for a finite value of M, one generally has

$$\widehat{B}_{\mathrm{RB},M} \neq B_{\mathrm{RB}}.$$

The conditional variance of the Monte Carlo approximation is

$$\boxed{\mathrm{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) = \frac{1}{M}\,\mathrm{Var}(B \mid S).}$$

Thus, the additional conditional simulation variance decreases inversely with the number of Stiefel configurations.

When the Stiefel randomization is included in the probability model and the conditional simulations are independent, the law of total variance gives

$$\begin{aligned}\mathrm{Var}(\widehat{B}_{\mathrm{RB},M}) &= \mathrm{Var}\{\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S)\} \\ &\quad + \mathbb{E}\{\mathrm{Var}(\widehat{B}_{\mathrm{RB},M} \mid S)\} \\ &= \mathrm{Var}(B_{\mathrm{RB}}) + \frac{1}{M}\mathbb{E}\{\mathrm{Var}(B \mid S)\}.\end{aligned}$$

Since

$$\mathrm{Var}(B) = \mathrm{Var}(B_{\mathrm{RB}}) + \mathbb{E}\{\mathrm{Var}(B \mid S)\},$$

we obtain

$$\boxed{\mathrm{Var}(\widehat{B}_{\mathrm{RB},M}) = \mathrm{Var}(B_{\mathrm{RB}}) + \frac{\mathrm{Var}(B) - \mathrm{Var}(B_{\mathrm{RB}})}{M}.}$$

This expression demonstrates that the unconditional variance of the finite-simulation estimator approaches the variance of the exact Rao–Blackwell estimator as M tends to infinity.

Moreover, conditional on S, the strong law of large numbers yields

$$\widehat{B}_{\mathrm{RB},M} \longrightarrow B_{\mathrm{RB}} \qquad \text{almost surely as } M \to \infty.$$

The corresponding mean-square convergence follows from

$$\mathbb{E}\left[(\widehat{B}_{\mathrm{RB},M} - B_{\mathrm{RB}})^2\right] = \frac{1}{M}\mathbb{E}\{\mathrm{Var}(B \mid S)\} \longrightarrow 0.$$

Consequently, the exact UMVUE property established for B_{RB} should not be transferred without qualification to a particular finite Monte Carlo realization. The finite-simulation estimator is an unbiased randomized approximation to the exact conditional estimator, provided that the underlying original estimator is unbiased and the conditional simulation is performed according to the correct parameter-free distribution.

An additional approximation arises in the present numerical implementation because the projection integral is evaluated using a finite number L of projection-direction pairs. Thus, the numerical procedure involves two distinct approximation mechanisms: the projection integration approximation controlled by L, and the conditional Stiefel integration approximation controlled by M. The theoretical variance identity above applies directly to the finite-L statistic when the projection directions are held fixed and the Stiefel simulations are independently generated. The additional projection-integration error must be considered separately when interpreting the numerical estimates relative to the exact population parameter.

12 Discussion and Limitations

The preceding development provides a conditional estimation framework for the projection-based multivariate Bergsma–Dassios dependence measure. The proposed construction combines a fourth-order U-statistic with conditional expectation under an unrestricted multivariate normal model. Its principal objective is to obtain an estimator that preserves unbiasedness while reducing sampling variance through conditioning on a complete sufficient statistic.

The theoretical development and numerical investigation address complementary aspects of the proposed framework. The theoretical results establish the unbiasedness and Rao–Blackwell properties under the specified statistical model, whereas the simulation study examines the numerical implementation of the conditional Stiefel representation and the finite-simulation performance of the resulting estimator.

12.1 Nature of the Multivariate Dependence Parameter

The dependence parameter considered in this work is defined through the averaging of univariate Bergsma–Dassios sign-kernel products over pairs of projection directions. Specifically,

$$\tau^*_{p,q} = \mathbb{E}\left[h(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4)\right],$$

where h incorporates integration over the projection spheres $\mathbb{S}^{p-1}$ and $\mathbb{S}^{q-1}$. This construction defines a particular projection-based multivariate extension of the original univariate-pair dependence measure.

It is therefore important to distinguish the parameter introduced here from other possible multivariate generalizations of Bergsma–Dassios τ^*. The theoretical properties established in this work apply to the particular kernel and population functional specified in the preceding sections. They should not automatically be attributed to alternative multivariate extensions based on different projection schemes, kernels, or aggregation procedures.

Furthermore, the numerical implementation approximates the projection-integrated kernel using a finite number L of projection-direction pairs. Accordingly, the simulation results concern the finite-L numerical version of the proposed construction. The identification of its limiting target with the intended population parameter requires the corresponding projection-integration normalization and convergence arguments to be established.

12.2 Unbiasedness and U-Statistic Structure

The proposed estimator

$$B = \widehat{\tau}^*_{p,q}$$

is a symmetric U-statistic of order four. Under iid sampling and the integrability of the kernel, it is unbiased for the specified population dependence parameter:

$$\mathbb{E}(B) = \tau^*_{p,q}.$$

In the present construction, the kernel is bounded, and hence the required integrability condition is satisfied.

The variance of the fourth-order U-statistic admits the classical Hoeffding decomposition

$$\mathrm{Var}(B) = \sum_{c=1}^{4} \frac{\binom{4}{c}^2}{\binom{n}{c}} \sigma_c^2,$$

where σ_c^2 denotes the variance of the cth Hoeffding projection component, under the adopted normalization.

This representation separates the sampling variability of the estimator into components associated with different orders of projection. It also provides a theoretical basis for investigating the asymptotic behaviour of the original U-statistic. The variance structure of the U-statistic, however, should be distinguished from the variance reduction obtained through Rao–Blackwellization. The former describes the sampling variability of the original estimator, whereas the latter results from conditioning on the complete sufficient statistic.

12.3 Dependence on the Multivariate Normal Model

The UMVUE characterization is established under the unrestricted multivariate normal family

$$\mathbf{Z}_i \overset{\text{iid}}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where both the mean vector $\boldsymbol{\mu}$ and the positive-definite covariance matrix $\boldsymbol{\Sigma}$ are unknown.

Under this model, the sample mean and sample scatter matrix constitute a complete sufficient statistic:

$$S = (\overline{\mathbf{Z}}, \mathbf{C}),$$

where

$$\mathbf{C} = \sum_{i=1}^{n} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^T.$$

The conditional distribution of the centered sample configuration given S is parameter-free and is represented by the uniform invariant distribution on the constrained Stiefel manifold.

Consequently,

$$B_{\text{RB}} = \mathbb{E}(B \mid S)$$

is an unbiased function of a complete sufficient statistic. By the Lehmann–Scheffé theorem, it is the unique UMVUE of $\tau_{p,q}^*$ within the specified unrestricted multivariate normal family, provided that the original estimator is unbiased for this parameter.

This conclusion is model-dependent. In particular, the completeness and sufficiency arguments do not establish a UMVUE under arbitrary nonparametric distributions. If the underlying population distribution is not multivariate normal, the conditional distribution of the sample given its mean and scatter matrix need not possess the same parameter-free Stiefel representation.

12.4 Role of the Nonparametric Hotelling Statistic

The nonparametric Hotelling-type statistic

$$A = T_{\text{NP}}^2$$

provides an alternative conditioning statistic. The corresponding conditional estimator is

$$B_A = \mathbb{E}(B \mid A).$$

Whenever B is unbiased and integrable, the law of iterated expectation gives

$$\mathbb{E}(B_A) = \mathbb{E}\{\mathbb{E}(B \mid A)\} = \mathbb{E}(B) = \tau_{p,q}^*.$$

Furthermore, the conditional variance identity implies

$$\mathrm{Var}(B_A) \le \mathrm{Var}(B).$$

However, this variance inequality alone does not establish that B_A is an implementable estimator. If the conditional distribution of B given A depends on unknown parameters, then the conditional expectation may also depend on unknown parameters.

A stronger comparison with the Rao–Blackwell estimator is available when the nonparametric Hotelling statistic is a measurable function of the complete sufficient statistic, that is,

$$A = \phi(S)$$

for some measurable function ϕ. Under this condition,

$$\sigma(A) \subseteq \sigma(S),$$

and the conditional variance decomposition yields

$$\boxed{\mathrm{Var}(B_{\mathrm{RB}}) \le \mathrm{Var}(B_A) \le \mathrm{Var}(B).}$$

The measurability condition is essential. If the particular nonparametric Hotelling statistic is not a function of the complete sufficient statistic, the above variance ordering between the two conditional estimators does not follow in general. Thus, an explicit mathematical definition of T^2_{NP} is necessary before this comparison can be claimed for a particular nonparametric Hotelling construction.

12.5 Computational Limitations

Although the conditional Stiefel representation provides an exact expression for the Rao–Blackwell estimator, its direct evaluation may be computationally demanding. The numerical difficulty arises from several sources. First, the dimension of the constrained Stiefel manifold increases with the sample size and the dimension of the joint observation vector. Second, each evaluation of the fourth-order U-statistic involves $\binom{n}{4}$ subsets of observations. Third, the multivariate kernel itself requires integration over two projection spheres.

The proposed Monte Carlo procedure addresses the integration over the constrained Stiefel manifold by replacing the exact conditional expectation with an average over randomly generated orthonormal configurations. For a finite number M of simulations, the approximation satisfies

$$\mathbb{E}(\widehat{B}_{\mathrm{RB},M} \mid S) = B_{\mathrm{RB}}$$

and

$$\mathrm{Var}(\widehat{B}_{\mathrm{RB},M} \mid S) = \frac{1}{M}\,\mathrm{Var}(B \mid S).$$

Thus, the additional conditional simulation variance decreases as M increases. Nevertheless, the Monte Carlo approximation is not identical to the exact Rao–Blackwell estimator for finite M.

The present implementation also approximates the projection-sphere integral using L random projection-direction pairs. The sensitivity analysis indicates that the resulting numerical estimates can vary with the choice of L, particularly when the number of projection directions is relatively small. Therefore, the projection-integration error and the conditional Stiefel-integration error should be investigated separately.

Furthermore, the total computational cost may remain substantial when n, p, or q is large.

Efficient implementation will therefore require attention to parallel computation, numerical integration over the projection spheres, and possible approximations to the fourth-order summation. The use of nested projection-direction collections and independent repeated direction sets would also provide a more controlled assessment of projection integration error.

12.6 Interpretation of the Dependence Parameter

Finally, the proposed multivariate dependence parameter should be distinguished from the cross-covariance matrix $\mathbf{\Sigma}_{XY}$. The cross-covariance matrix characterizes second-order linear association between $\mathbf{X}$ and $\mathbf{Y}$. In contrast, $\tau^*_{p,q}$ is defined through a projection-based four-observation sign-kernel construction.

Although both quantities describe aspects of the relationship between the two random vectors, they are mathematically different population functionals. The proposed estimator therefore targets $\tau^*_{p,q}$ and should not be interpreted as an estimator of $\mathbf{\Sigma}_{XY}$.

13 Conclusion

A conditional estimation framework has been developed for a projection-based multivariate Bergsma–Dassios dependence measure. The proposed estimator is a fourth-order U-statistic and is unbiased for the corresponding population dependence parameter under the stated sampling assumptions. Under the unrestricted multivariate normal model, the sample mean and sample scatter matrix form a complete sufficient statistic. Conditioning the original estimator on this statistic yields a parameter-free Rao–Blackwell estimator, which is established as the unique UMVUE through the Lehmann–Scheffé theorem.

The conditional Stiefel representation provides an exact mathematical characterization of this conditional estimator and a computational route for approximating it. The numerical investigation verified the orthogonality and centering restrictions of the generated Stiefel matrices and confirmed that the reconstructed samples preserve the observed sample mean and scatter matrix to numerical precision.

The repeated-sampling experiment indicated that the Monte Carlo Rao–Blackwell approximation achieved lower empirical variance, RMSE, and MAE than the original U-statistic for the considered simulation configuration. In particular, the observed empirical variance reduction was approximately 48%, with an estimated relative efficiency of approximately 1.9231. These findings provide preliminary numerical evidence for the variance-reduction potential of the proposed conditional estimation procedure.

The numerical investigation also highlighted the importance of controlling the two sources of Monte Carlo approximation error: projection integration and conditional Stiefel integration. The finite-simulation estimator approaches the exact conditional estimator as the number of Stiefel simulations increases, while more projection directions are required to obtain a stable approximation to the projection-integrated kernel.

The proposed framework therefore provides a theoretical and computational foundation for studying conditional variance reduction in multivariate nonparametric dependence estimation. Further investigation is required to develop computationally efficient integration procedures, examine larger sample sizes and higher-dimensional settings, and assess the finite-sample performance of the method under a broader range of dependence structures.

Acknowledgements

The author thanks the colleagues for their helpful comments and suggestions to develop and improve this manuscript.

Funding

No external funding was received for this study.

Conflict of Interest

The author declares that there are no conflicts of interest.

Supplementary Files

Visit `https://github.com/sthdas999/Hotelling-T-square-Given-Tau-star`.

References

[1] Wicher Bergsma and Angelos Dassios. A consistent test of independence based on a sign covariance related to kendall's tau. *Bernoulli*, 20(2):1006–1028, 2014.

[2] Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.

[3] Erich L. Lehmann and George Casella. *Theory of Point Estimation*. Springer, New York, 2 edition, 1998.

[4] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RCode1.pdf
- OutputR.txt