Rao's U-Statistic in Discriminant Analysis: Theory, Admissibility, and Applications

Abstract

C. R. Rao's contributions to multivariate analysis are among the most influential developments in modern statistics. This report provides an in-depth analysis of Rao's U-statistic, which is used to test the additional information hypothesis in two-group discriminant analysis. We examine the theoretical foundations of the statistic, the equivalent conditions for variable redundancy, the occurrence of Rao's Paradox under high dimensionality, and the rigorous decision-theoretic proofs of the admissibility of Rao's U-test over Hotelling's T^2 under restricted alternatives. Finally, we discuss the role of the U-statistic in modern variable selection via information criteria such as Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). We also examine how these classical concepts are extended into reproducing kernel Hilbert spaces (RKHS) for non-linear multi-class classification.

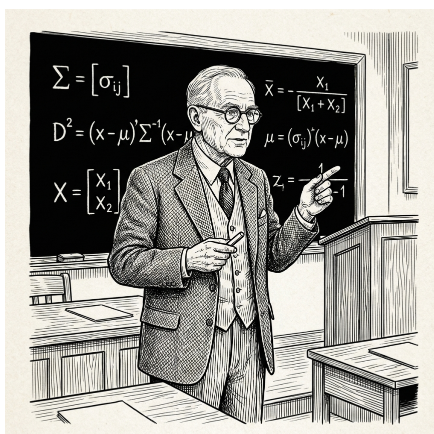

Figure 1: Historical scholarly portrait of a mid-20th-century statistician, representing the era of C. R. Rao's pioneering contributions at the Indian Statistical Institute and Cambridge.

1 Introduction and the Additional Information Hypothesis

In two-group discriminant analysis, we are interested in separating observations from two distinct populations, Π_1 and Π_2, based on a vector of $p+q$ measurements. Let $\mathbf{Y}$ be a $(p+q)$-variate random vector partitioned as:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} \tag{1}$$

where $\mathbf{Y}_1$ is a $p \times 1$ subvector representing the primary, already available variables, and $\mathbf{Y}_2$ is a $q \times 1$ subvector representing additional, secondary variables whose discriminative contribution we wish to evaluate.

We assume that the observations in each population follow a multivariate normal distribution:

$$\mathbf{Y} \mid \Pi_i \sim N_{p+q}(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma}), \quad i = 1, 2 \tag{2}$$

where the mean vectors $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ and the positive-definite covariance matrix $\boldsymbol{\Sigma}$ are partitioned conformably with $\mathbf{Y}$:

$$\boldsymbol{\mu}^{(i)} = \begin{bmatrix} \boldsymbol{\mu}_1^{(i)} \\ \boldsymbol{\mu}_2^{(i)} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \tag{3}$$

Rao (1946) proposed a rigorous framework to determine if the secondary subvector $\mathbf{Y}_2$ provides any statistically significant additional information for discriminating between the two populations in the presence of $\mathbf{Y}_1$. This is formulated as the additional information hypothesis:

$$H_0 : \delta^2 = \delta_1^2 \quad \text{vs.} \quad H_1 : \delta^2 > \delta_1^2 \tag{4}$$

where δ^2 and δ_1^2 represent the population squared Mahalanobis distances based on the full vector $\mathbf{Y}$ and the primary subvector $\mathbf{Y}_1$, respectively:

$$\delta^2 = (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) \tag{5}$$

$$\delta_1^2 = (\boldsymbol{\mu}_1^{(1)} - \boldsymbol{\mu}_1^{(2)})' \boldsymbol{\Sigma}_{11}^{-1} (\boldsymbol{\mu}_1^{(1)} - \boldsymbol{\mu}_1^{(2)}) \tag{6}$$

1.1 The Redundancy Equivalence Theorem

Rao (1970) established that the additional variables $\mathbf{Y}_2$ are redundant (i.e., do not contribute to group separation) if and only if any of the following three conditions hold:

1. **Distance Equality**: The population squared Mahalanobis distances are equal: $\delta^2 = \delta_1^2$.
2. **Conditional Mean Equality**: The conditional mean vectors of $\mathbf{Y}_2$ given $\mathbf{Y}_1$ are identical in both populations:
$$\boldsymbol{\mu}_{2\cdot 1}^{(1)} = \boldsymbol{\mu}_{2\cdot 1}^{(2)} \tag{7}$$
where $\boldsymbol{\mu}_{2\cdot 1}^{(i)} = \boldsymbol{\mu}_2^{(i)} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1^{(i)}$ for $i = 1, 2$.
3. **Coefficient Vanishing**: The coefficients $\boldsymbol{\beta}_2$ corresponding to $\mathbf{Y}_2$ in the population linear discriminant function $\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}) = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ are zero:
$$\boldsymbol{\beta}_2 = \mathbf{0} \tag{8}$$

These equivalences provide a unified bridge between classification theory, conditional probability distributions, and the geometry of multi-dimensional spaces.

2 Rao's U-Statistic and Its Exact Distribution

Let $\mathbf{S}$ be the pooled sample covariance matrix with $n = n_1 + n_2 - 2$ degrees of freedom based on independent samples of sizes n_1 and n_2 from Π_1 and Π_2. Let D_{p+q}^2 and D_p^2 represent the sample squared Mahalanobis distances:

$$D_{p+q}^2 = (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)})' \mathbf{S}^{-1} (\bar{\mathbf{x}}^{(1)} - \bar{\mathbf{x}}^{(2)}) \tag{9}$$

$$D_p^2 = (\bar{\mathbf{x}}_1^{(1)} - \bar{\mathbf{x}}_1^{(2)})' \mathbf{S}_{11}^{-1} (\bar{\mathbf{x}}_1^{(1)} - \bar{\mathbf{x}}_1^{(2)}) \tag{10}$$

where $\bar{\mathbf{x}}^{(i)}$ represents the sample mean vectors. To test the hypothesis $H_0 : \delta^2 = \delta_1^2$, Rao proposed the U-test statistic:

$$U = \frac{n-p-q-1}{q} \cdot \frac{n_1 n_2 (D_{p+q}^2 - D_p^2)}{n(n-2) + n_1 n_2 D_p^2} \tag{11}$$

whose exact null distribution is the classical F-distribution:

$$U \sim F_{q,\, n-p-q-1} \tag{12}$$

Alternatively, the statistic can be expressed in terms of the ratio of Hotelling's T^2 statistics under subvector partitioning:

$$U_{\text{ratio}} = \frac{1 + \frac{T_p^2}{n}}{1 + \frac{T_{p+q}^2}{n}} \tag{13}$$

where $T_p^2 = \frac{n_1 n_2}{n_1+n_2} D_p^2$ and $T_{p+q}^2 = \frac{n_1 n_2}{n_1+n_2} D_{p+q}^2$. Here, U_{ratio} acts as a monotonic transformation of the conditional Wilks' Lambda ($\Lambda_{2 \cdot 1}$), and the test is conducted using the exact F-transformation:

$$\frac{n-p-q-1}{q} \left[\frac{1}{U_{\text{ratio}}} - 1 \right] \sim F_{q,\, n-p-q-1} \tag{14}$$

3 Rao's Paradox and the Curse of Dimensionality

A fundamental issue in multivariate statistical inference is that adding more variables does not always increase the power of a statistical test. First named by Healy (1969) and explained in detail by Rencher (2002), Rao's Paradox is an extreme manifestation of this principle.

3.1 The Phenomenon

Rao's Paradox occurs when:

1. Separate univariate two-sample t-tests on two individual variables X_1 and X_2 are both highly significant.
2. Yet, a joint multivariate Hotelling's T^2 test on the bivariate vector $[X_1, X_2]'$ fails to show any significant difference between the populations.

Conversely, the opposite paradox can occur: neither individual variable is significant on its own, but their joint combination is highly significant due to the correlation structure.

3.2 Geometric Interpretation

As illustrated in Figure 2, the rejection region of the marginal t-tests is the area outside the square $[-1.96, 1.96] \times [-1.96, 1.96]$. The rejection region of Hotelling's T^2 is the area outside the critical ellipse of the Mahalanobis distance.

When the variables are correlated ($\rho = 0.5$ in Figure 2), the ellipse is slanted. In the upper-right and lower-left regions, the ellipse extends beyond the corners of the box. A sample falling in these regions lies outside the box (meaning individual t-tests reject H_0) but inside the ellipse (meaning the joint Hotelling's T^2 test fails to reject H_0).

This dilution of power occurs because the joint multivariate test must penalize the critical boundary for the added dimension. If the physical distance added by the second variable is small relative to the noise, the joint test's power is crushed by the "curse of dimensionality."

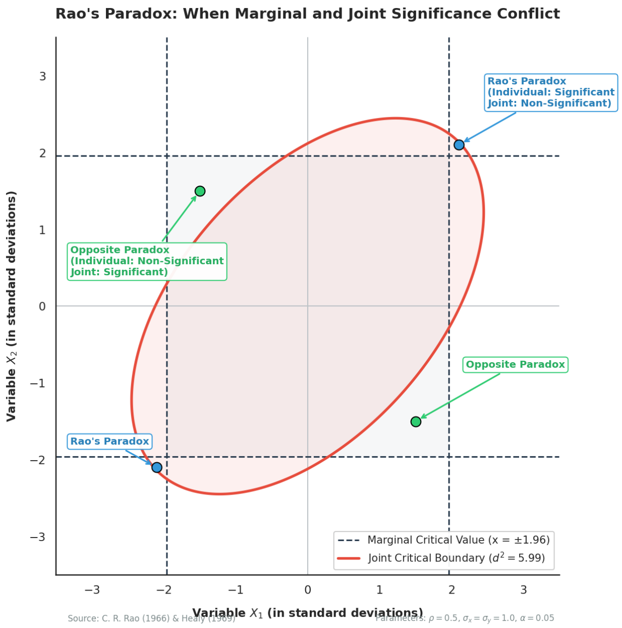

Figure 2: Geometric representation of Rao's Paradox (Healy 1969). The vertical and horizontal lines represent individual marginal critical boundaries (± 1.96 standard deviations) at $\alpha = 0.05$. The shaded region represents the joint multivariate critical region for Hotelling's T^2 ($d^2 = 5.99$). Points falling in the top-right and bottom-left quadrants outside the box but inside the ellipse illustrate the paradox, while the opposite paradox occurs in the top-left and bottom-right corners.

4 Admissibility of Invariant Tests

A long-standing debate in multivariate theory has been the choice between Hotelling's T^2 and Rao's U-test for testing subvectors. Marden and Perlman (1980) proved that Hotelling's T^2-test is admissible under a global alternative but claimed that Rao's U-test is inadmissible under their set-up.

Tsai (2017) resolved this apparent contradiction by showing that the admissibility depends entirely on the parameter space of the alternative hypothesis. When we test against a restricted alternative that incorporates the prior covariate information $\boldsymbol{\theta}_2 = \mathbf{0}$, Rao's U-test is admissible, whereas Hotelling's T^2 becomes inadmissible.

4.1 Convexity of the Acceptance Region

Under the group G of nonsingular lower-triangular block matrices, the acceptance region of Rao's U-test can be written on the space $(\bar{\mathbf{X}}, \mathbf{S})$ as:

$$A_U = \{(\bar{\mathbf{X}}, \mathbf{S}) \mid n(n-1)\bar{\mathbf{X}}'\mathbf{B}^+(\mathbf{S})\bar{\mathbf{X}} \leq k\} \tag{15}$$

where $\mathbf{B}^+(\mathbf{S})$ is the Moore-Penrose generalized inverse of the partitioned covariance matrix:

$$\mathbf{B}^+(\mathbf{S}) = \mathbf{S}^{-1} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22}^{-1} \end{bmatrix} \tag{16}$$

Tsai proved that $\mathbf{B}^+(\mathbf{S})$ is a matrix-convex function on the set of positive-definite matrices $\mathbf{S}$. By utilizing this property and the Birnbaum-Stein theorem, the acceptance region A_U is shown to be a closed, convex set in $\mathbb{R}^p \times \mathbf{S}$. Since the acceptance region is convex and disjoint with the half-space defined by the likelihood ratio under the restricted alternative, Rao's U-test is proven to be α-admissible (Tsai 2017).

4.2 Inadmissibility of Hotelling's T^2

Eaton (1970) established a complete class theorem for one-sided and restricted alternatives. Let $\Omega_1 = \{\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta} \mid \boldsymbol{\theta}_1 \neq \mathbf{0}, \boldsymbol{\theta}_2 = \mathbf{0}\}$. Let $V \subset \mathbb{R}^p$ be the smallest closed convex cone containing Ω_1, and let V^- be its dual cone:

$$V^- = \{\mathbf{w} \mid \mathbf{x}'\mathbf{w} \leq 0, \ \forall \mathbf{x} \in V\} \tag{17}$$

For any invariant test to be admissible, its acceptance region for a fixed sample covariance matrix $\mathbf{S}$ must contain the dual cone V^-.

Since $V^- \cong \mathbb{R}^q$ is an unbounded set, whereas the acceptance region of Hotelling's T^2 is a bounded ellipsoid in $\mathbb{R}^p$ for a fixed $\mathbf{S}$, the acceptance region fails to contain the dual cone. Therefore, Hotelling's T^2-test is inadmissible under the restricted alternative because it completely ignores the covariate constraint $\boldsymbol{\theta}_2 = \mathbf{0}$ (Tsai 2017).

5 Modern Applications in Variable Selection

In modern high-dimensional data analysis, the mathematical foundation of Rao's U-statistic is used extensively for feature selection and model building. Rather than performing multiple pairwise hypothesis tests (which inflates the family-wise error rate), modern procedures use information criteria derived directly from the U-statistic's likelihood ratio representation.

5.1 Akaike's Information Difference (A_j)

Fujikoshi (1985) derived the Akaike Information Criterion (AIC) for a candidate variable selection model M_j which assumes that only a subset of variables $\mathbf{Y}_j$ (containing p_j elements) has non-zero discriminant coefficients. The AIC difference between the subset model and the full model M_ω (containing p variables) is given by:

$$A_j = \mathrm{AIC}_j - \mathrm{AIC}_\omega = n \log \left[1 + \frac{g^2(D_\omega^2 - D_j^2)}{n - 2 + g^2 D_j^2} \right] - 2(p - p_j) \tag{18}$$

where $g = \sqrt{n_1 n_2 / n}$. Similarly, the BIC difference is obtained by replacing the multiplier 2 with $\log n$. The optimal model is selected by finding the subset j that minimizes these differences.

These criteria have been shown to possess strong consistency properties under both large-sample ($n \to \infty$) and high-dimensional ($n/p \to c \in (0, 1)$) frameworks, providing a rigorous and computationally efficient alternative to stepwise selection algorithms (Fujikoshi and Sakurai 2019).

5.2 Non-linear Kernel Generalizations: Multi-Class GDA

The classical linear discriminant analysis (LDA) and Rao's U-test are fundamentally restricted to linear decision boundaries. To classify non-linearly separable data, researchers have developed Generalized Discriminant Analysis (GDA) using kernel methods. GDA maps input patterns $\mathbf{x}$ into a high-dimensional reproducing kernel Hilbert space (RKHS) $\mathcal{F}$ via a non-linear mapping $\boldsymbol{\phi} : \mathbb{R}^d \to \mathcal{F}$, where linear discrimination is then performed.

However, standard GDA becomes computationally intractable for large datasets because it requires the manipulation of the $N \times N$ Gram matrix $\mathbf{K}$. To overcome this, Abdallah, Richard, and Lengeallé proposed a sequential approach based on a gradient descent strategy that completely avoids the storage or inversion of the large Gram matrix.

Figures 3 and 4 demonstrate the empirical performance of this sequential GDA approach on non-linear multi-class datasets.

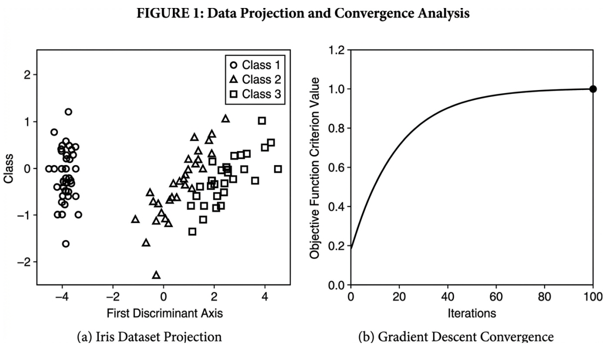

Figure 3: Sequential Generalized Discriminant Analysis (GDA) on the Iris dataset. Left (a): Projection of each class (50 elements) on the first sequential discriminant axis, illustrating clear non-linear separability. Right (b): Convergence curve showing the GDA criterion value smoothly rising to its maximum as a function of iteration count.

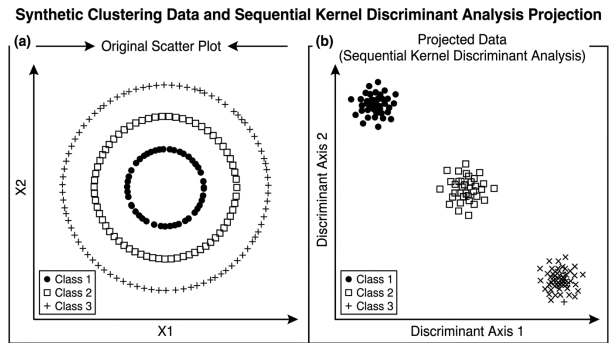

Figure 4: Sequential GDA on synthetic non-linearly separable data consisting of three concentric classes. Left (a): Scatter plot of the raw concentric circles. Right (b): Bivariate projection of the whole dataset onto the first two sequential discriminant axes, illustrating how kernel GDA unwinds non-linear geometric structures into perfectly linear separable clusters.

6 Conclusion

Rao's U-statistic remains a cornerstone of multivariate statistical theory. It elegantly addresses the additional information hypothesis, provides a warning against the curse of dimensionality through Rao's Paradox, and showcases the necessity of restricted parameter spaces in decision-theoretic admissibility. Whether in classical anthropology or modern high-dimensional machine learning, the principles established by C. R. Rao continue to guide robust multivariate feature selection and classifier design.

References

[1] Abdallah, F., Richard, C., and Lengeallé, R. (2003). A sequential approach for multi-class discriminant analysis with kernels. LM2S Laboratoire, Université de Technologie de Troyes.

[2] Eaton, M. L. (1970). A complete class theorem for multidimensional one-sided alternatives. The Annals of Mathematical Statistics, 41(6), 1884-1888.

[3] Fujikoshi, Y. (1985). Selection of variables in discriminant analysis and canonical correlation analysis. Journal of Multivariate Analysis, 17, 27-37.

[4] Fujikoshi, Y., and Sakurai, T. (2019). Consistency of test-based method for selection of variables in high-dimensional two-group discriminant analysis. Japanese Journal of Statistics and Data Science, 2, 155-171.

[5] Gupta, A. K., Xu, J., and Fujikoshi, Y. (2006). An asymptotic expansion of the distribution of Rao's U-statistic under a general condition. Journal of Multivariate Analysis, 97(2), 492-513.

[6] Healy, M. J. R. (1969). Rao's paradox concerning multivariate tests of significance. Biometrics, 25(2), 411-413.

[7] Marden, J., and Perlman, M. D. (1980). Invariant tests for means with covariates. The Annals of Statistics, 8(1), 25-63.

[8] Rao, C. R. (1946). Tests with discriminant functions in multivariate analysis. Sankhyā, 7, 407-414.

[9] Rao, C. R. (1970). Inference on discriminant function coefficients. Essays in Probability and Statistics, 587-602.

[10] Tsai, M. T. (2017). Admissibility of invariant tests for means with covariates. Journal of Modern Applied Statistical Methods, 16(2), 486-488.