ISSN 2309-4001 e-ISSN 3083-7545 https://bmj.chnu.edu.ua

Bukovinian Math. Journal 2026, **14** (2) , 37–54 Буковинський матем. журнал 2026, Т.14, №2, 37–54

DOI: 10.31861/bmj2026.02.04

Holdovanskyi Vladyslav [1D], Alieksieiev Vladyslav [1D]

A DETERMINATION-BASED CORRELATION COEFFICIENT FOR MULTIVALUED DEPENDENCIES VIA BAYESIAN MIXTURE MODELS

Measuring statistical dependence between two random variables is a fundamental problem of data analysis, yet no single measure captures every type of relationship while remaining interpretable. Classical coefficients detect only narrow classes of dependence: linear for Pearson, monotonic for Spearman. More general measures are either computationally expensive, prone to overestimation, or uninformative about the structure of the relationship. This paper makes two contributions. First, we propose a unified framework for comparing dependence measures of different mathematical natures, based on the functional class of a coefficient and a set of linking properties anchored to the Pearson correlation as a common reference. Second, building on the correlation ratio, we develop a determination-based correlation coefficient (DBCC) that captures multivalued dependencies, in which a single value of the predictor may correspond to several states of the response. The coefficient decomposes the conditional distribution within local bins through a soft-assignment Bayesian Gaussian mixture model, replacing the discontinuous mode-seeking of an earlier kernel-based realisation. We prove the consistency of the estimator via an oracle decomposition that separates the algorithmic and statistical components of the error, and we show that the coefficient is computed in linear time. The construction retains the transparent variance-explained interpretation of the classical coefficient of determination R^2.

Key words and phrases: dependence measure, correlation coefficient, coefficient of determination, multivalued dependence, Bayesian mixture model, nonlinear correlation, consistency.

Lviv Polytechnic National University, Lviv, Ukraine
e-mail: *vladyslav.a.holdovanskyi@lpnu.ua (Holdovanskyi V.A.), vladyslav.i.alieksieiev@lpnu.ua (Alieksieiev V.I.)*

Introduction

Measuring statistical dependence between two random variables remains one of the most fundamental problems in data analysis. Although the Pearson correlation coefficient was proposed more than a century ago and many alternatives have since appeared, no single standardised and universal measure exists.

The obstacle lies in the mathematical limitations of existing measures. Classical criteria capture only specific relationships: the Pearson coefficient is effective only for linear dependence and can vanish in the presence of pronounced nonlinear relationships, while Spearman's rank correlation is limited to monotonic dependence. More sophisticated measures have their own drawbacks. The Randomized Dependence Coefficient (RDC) can overestimate the strength of the relationship on

UDC 519.237.5
2020 *Mathematics Subject Classification:* 62H20, 62G05, 62F12.

© Holdovanskyi V.A., Alieksieiev V.I., 2026

certain distributions; distance correlation has complexity $\mathcal{O}(N^2)$, which hinders its use on large datasets; and the exact maximal information coefficient (MIC) is computationally expensive in its grid-resolution parameter.

Moreover, a single dependence measure, even one universally capable of detecting any relationship, is analytically uninformative: a scalar value reports the presence and strength of the dependence but not its nature - direction, type of nonlinearity, multivaluedness, or heteroscedasticity. Modern practice therefore computes several measures jointly. In principle their comparison reveals the structure of the data, but each criterion rests on its own concept and nonlinear scale, so a direct comparison of values is more a heuristic than a rigorous procedure.

Given these gaps, this work pursues two objectives. The first is a more formalised approach to the joint analysis of dependence measures. Although partly heuristic, the proposed framework offers a systematic way to juxtapose coefficients with different mathematical foundations, built around the functional class of a coefficient and a set of linking properties anchored to the Pearson correlation as a common reference.

The second and central objective is the introduction of a determination-based correlation coefficient (DBCC), a generalisation of the coefficient of determination. DBCC is adapted to multivalued dependencies, in which a single value of X may correspond to several states of Y (closed curves, ellipses, hidden categorical structure), while retaining the transparent interpretation of the classical R^2. The coefficient decomposes the conditional distribution of Y within local intervals of X through a soft-assignment Bayesian Gaussian mixture model, replacing an earlier kernel-based mode-seeking realisation whose discontinuous response and slow nonparametric rate complicated both practical use and asymptotic analysis.

We prove the consistency of the DBCC estimator by an oracle decomposition that separates the algorithmic error (the local mixture fits) from the statistical error (averages of identically distributed quantities), and we establish that the coefficient is computed in linear time $\mathcal{O}(N)$. We also analyse the behaviour of the measure under extreme outliers, identifying the L_2 origin of its sensitivity and outlining a robust variant as future work.

The remainder of the paper is organised as follows. Section 1 reviews related work. Section 2 develops the unified comparison framework. Section 3 analyses the correlation ratio as the base coefficient. Section 4 presents the initial kernel-based realisation and its limitations. Section 5 introduces the soft-assignment Bayesian mixture formulation and positions DBCC within the framework. Section 6 establishes the asymptotic theory. Section 7 analyses robustness, and Section 8 treats computational complexity. Section 9 reports experiments and Section 10 concludes.

1 Related Works

The study of dependence measures begins with the Pearson correlation coefficient [9], defined as normalised covariance, and Spearman's rank correlation [18], which applies the same logic to ranks to capture monotonic dependence. The search for a universal measure led to the maximal correlation of Hirschfeld, Gebelein and Rényi [5,6,13], and to the axiomatic characterisations of an "ideal" measure proposed by Rényi [13] and, for concordance, by Scarsini [16].

Modern nonparametric measures include distance correlation [19], which detects arbitrary dependence through distances between observations; the maximal information coefficient [14], based on adaptive grid partitions; the Randomized Dependence Coefficient [7], which combines random nonlinear projections with canonical correlation; and Chatterjee's rank coefficient [3], an asymmetric measure that equals one if and only if Y is a function of X. The present work builds on the correlation ratio, whose theoretical properties trace back to Rényi [13], and on the theory of finite mixture models, whose identifiability [20], maximum-likelihood consistency [12], and overfitted-prior behaviour [15] underpin our asymptotic analysis.

2 A Unified Framework for Comparing Dependence Measures

Before formalising the proposed coefficient, we review how the field has converged on the requirements such a measure should satisfy. The Pearson coefficient [9], geometrically the cosine of the angle between centred vectors, is sensitive to outliers and cannot detect nonlinear dependence. Applying Pearson's logic to ranks, Spearman [18] produced a measure of monotonic dependence. Subsequent work moved towards generalisation through axiomatic systems characterising an "ideal" dependence measure.

Rényi [13] first formalised the requirements for a universal dependence measure $\delta(X, Y)$, proposing seven axioms: domain (defined for any non-constant random variables); symmetry $\delta(X, Y) = \delta(Y, X)$; normalisation $0 \leq \delta(X, Y) \leq 1$; independence criterion $\delta(X, Y) = 0$ if and only if $X \perp Y$; strict-dependence criterion $\delta(X, Y) = 1$ if and only if there is exact functional dependence almost surely; invariance under strictly monotonic transformations of either variable; and consistency with the bivariate normal case, where $\delta(X, Y) = |\rho|$. Pearson's coefficient violates the independence and strict-dependence criteria. Rényi proved that the Hirschfeld–Gebelein–Rényi maximal correlation [5,6] satisfies all seven axioms, but in practice this measure is computationally complex, prone to false overestimation on finite samples, and - because of its maximal generality - conveys nothing about the shape of the dependence.

A separate line of work concerns concordance. Scarsini [16] formalised the axioms for a measure of concordance $\kappa(X, Y)$ based on copulas: range $-1 \leq \kappa \leq 1$; symmetry; independence implies $\kappa = 0$; sign antisymmetry; pointwise monotonicity in copulas; and $\kappa = \pm 1$ if and only if Y is an almost-sure strictly monotone function of X. Spearman's and Kendall's coefficients satisfy these conditions but cannot detect non-monotonic dependence.

The most relevant framework for our purposes is the directional one. In real systems, dependence is rarely symmetric: X may completely determine Y while the converse fails, as for $Y = X^2$. A directed measure $D(X \mid Y)$ is required to satisfy normalisation, an independence criterion, a complete-predictability criterion ($D(X \mid Y) = 1$ if and only if $X = f(Y)$ almost surely), and invariance under strictly monotonic bijections. The symmetry axiom is dropped. This class includes the generalised coefficient of determination, normalised mutual information, and Chatterjee's correlation [3]. Each axiomatic system reflects a distinct facet of dependence, so only a joint analysis of coefficients of different natures can yield a comprehensive understanding of the structure of the dependence.

2.1 A Common Core and a Linking Reference

Despite their different natures, most criteria share a common "core" of properties: normalisation, the independence criterion, and the strict-dependence criterion; symmetry is optional. To enable systematic comparison, we combine these core properties with linking properties that anchor each coefficient to a common reference. We adopt the Pearson coefficient $\rho(X, Y)$ as that reference, since many measures (Spearman's coefficient, the correlation ratio, the maximal correlation) can be written as a supremum of Pearson correlation over a class of transformations.

Definition 2.1. *The* functional class *of a dependence measure ρ^*, denoted $\mathcal{T}_{\rho^*}$, is the set of pairs of transformations (f, g) for which ρ^* achieves its maximal informativeness.*

Example 2.1. *For Pearson's coefficient ρ, $\mathcal{T}_\rho$ consists of pairs of affine transformations. For Spearman's coefficient ρ_S, $\mathcal{T}_{\rho_S}$ consists of pairs of strictly monotonic transformations. For the maximal correlation $\rho^{\max}$, $\mathcal{T}_{\rho^{\max}}$ consists of all pairs of Borel-measurable transformations.*

Following the logic of Rényi's seventh axiom, we propose a unified set of properties against which any dependence measure $\rho^*(X, Y)$ should be evaluated to be comparable with others.

1. **Normalisation:** $0 \le |\rho^*(X,Y)| \le 1$; the range is $[-1,1]$ if ρ^* captures direction, otherwise $[0,1]$.

2. **Affine invariance:** $\rho^*(a_1X + b_1, a_2Y + b_2) = \rho^*(X,Y)$ for constants $a_1, a_2 \neq 0$ and b_1, b_2.

3. **Independence criterion:** $\rho^*(X,Y) = 0$ if and only if X and Y are stochastically independent.

4. **Strict-dependence criterion:** $|\rho^*(X,Y)| = 1$ if and only if there is exact functional dependence $Y = f(X)$ or $X = g(Y)$ with $(f,g) \in \mathcal{T}_{\rho^*}$.

5. **Linear consistency:** if (X,Y) is bivariate normal with Pearson correlation ρ, then $\rho^*(X,Y) = |\rho(X,Y)|$.

6. **Consistency in the general case:** if the true dependence is generated by $(f,g) \in \mathcal{T}_{\rho^*}$, then $\rho^*(X,Y) = \sup_{(f,g)\in\mathcal{T}_{\rho^*}} |\rho(f(X), g(Y))|$.

7. **Suboptimality outside the class:** if the true dependence is generated by $(f^*, g^*) \notin \mathcal{T}_{\rho^*}$, then $|\rho^*(X,Y)| < \sup_{(f,g)\in\mathcal{T}^+} |\rho(f(X), g(Y))|$ for some $\mathcal{T}^+ \supsetneq \mathcal{T}_{\rho^*}$.

Remark 2.1. *At this stage these properties form a heuristic framework rather than a fully proven axiomatic system. Even so, they guarantee macro-level comparability across coefficients of different natures, and the suboptimality criterion (Property 7) explains why certain coefficients underestimate complex dependence patterns, providing a starting point for a future theory of comparability.*

3 The Correlation Ratio as the Base Coefficient

The generalised coefficient is built on the classical correlation ratio

$$\eta^2(Y \mid X) = \frac{\mathrm{Var}(\mathbb{E}[Y \mid X])}{\mathrm{Var}(Y)} = 1 - \frac{\mathbb{E}[\mathrm{Var}(Y \mid X)]}{\mathrm{Var}(Y)}.$$

We use the residual-to-total variance form, $1 - \mathrm{SS}_{\mathrm{res}}/\mathrm{SS}_{\mathrm{tot}}$, so that the result is interpretable as a coefficient of determination.

Rényi [13] showed that $\eta(Y \mid X)$ satisfies most of his axioms, with three exceptions: it requires $\mathrm{Var}(Y) < \infty$ (and is undefined for the Cauchy distribution); it can vanish under a closed-curve dependence, because $\mathbb{E}[Y \mid X]$ is then constant; and it is not symmetric. In the directional framework of the previous section these are design features, not drawbacks. Rényi also proved

$$\eta(Y \mid X) = \sup_f |\rho(f(X), Y)|, \tag{1}$$

so the functional class is $\mathcal{T}_\eta = \{(f,g) : f \text{ Borel-measurable}, g \equiv Y\}$. The coefficient detects a wide range of nonlinear dependencies, provided Y is not subject to nonlinear distortion.

Remark 3.1. *Unlike distance correlation [19], with complexity $\mathcal{O}(N^2)$, or the maximal information coefficient [14], the correlation ratio admits an $\mathcal{O}(N)$ implementation and a direct interpretation as the proportion of variance of Y explained by the nonlinear effect of X. The contrast between $\eta^2(Y \mid X)$ and the maximal correlation reveals whether the relationship also requires a nonlinear transformation of Y.*

4 Initial Realisation: Kernel-Based Mode Seeking

The classical correlation ratio is restricted to single-valued mappings, relying on a single conditional expectation $\mathbb{E}[Y \mid X = x]$. It is therefore blind to topologies where one value of X corresponds to several states of Y, such as closed curves, ellipses, or hidden categorical structure.

In a preceding work the authors proposed an extension based on *kernel density estimation with local maxima search*. The domain of X was discretised into K bins B_k; for each bin a kernel density estimator of $p(y \mid x \in B_k)$ was constructed, and its local maxima were treated as branch centres. Each point was hard-assigned to its nearest mode, partitioning the bin into clusters $S_{k,j}$, and the residual variance was computed within each cluster:

$$\widehat{\mathrm{Var}}^{*}_{\mathrm{res}} = \sum_{k=1}^{K} P(X \in B_k) \sum_{j=1}^{c_k} w_{k,j} \,\mathrm{Var}(Y \mid X \in B_k, Y \in S_{k,j}).$$

Proof-of-concept experiments confirmed that this captures multi-branch topologies the classical correlation ratio cannot.

The kernel-based realisation has several structural weaknesses. The mapping from data to mode locations is discontinuous: as noise increases, modes merge abruptly, producing a discontinuous response of the coefficient and complicating asymptotic analysis. At branch crossings the conditional distribution becomes unimodal, so hard assignment misclassifies points and inflates local variance. The number of peaks depends strongly on the bandwidth h, and asymptotic guarantees require an awkward undersmoothing condition $Nh^4 \to 0$. Finally, kernel density estimation converges at the slow rate $\mathcal{O}(N^{-2/5})$, which propagates to the coefficient. These limitations motivate the soft-assignment formulation of the next section.

5 Soft-Assignment Bayesian Mixture Models

5.1 From Hard Partition to Soft Weights

Instead of partitioning points into branches, we model the conditional distribution within each bin B_k as a finite mixture

$$p(y \mid x \in B_k) = \sum_{j=1}^{c_k} \pi_{k,j}\, f(y \mid \theta_{k,j}),$$

with $\pi_{k,j} \geq 0$, $\sum_j \pi_{k,j} = 1$, and Gaussian components $f(\cdot \mid \theta)$ [8]. Posterior inference over $(\pi_{k,j}, \theta_{k,j})$ is performed by variational Bayesian methods, so that each observation (X_i, Y_i) with $X_i \in B_k$ carries posterior responsibilities

$$\gamma_{i,k,j} = \frac{\pi_{k,j} f(Y_i \mid \theta_{k,j})}{\sum_{l=1}^{c_k} \pi_{k,l} f(Y_i \mid \theta_{k,l})}, \qquad \sum_{j=1}^{c_k} \gamma_{i,k,j} = 1.$$

No point is forcibly assigned to a single branch. The within-component variance has a closed form: by the law of total variance,

$$\mathrm{Var}(Y \mid X \in B_k) = \underbrace{\sum_{j=1}^{c_k} \pi_{k,j} \sigma^2_{k,j}}_{\text{within-component}} + \underbrace{\sum_{j=1}^{c_k} \pi_{k,j} (\mu_{k,j} - \bar{\mu}_k)^2}_{\text{between-component}},$$

where $\bar{\mu}_k = \sum_j \pi_{k,j} \mu_{k,j}$. The first term is the residual variance, exactly what we need.

Definition 5.1. *Let X be discretised into bins $\{B_k\}_{k=1}^{K}$ and let $(\pi_{k,j}, \mu_{k,j}, \sigma_{k,j}^2)_{j=1}^{c_k}$ be the parameters of a finite Bayesian mixture fitted to $p(y \mid x \in B_k)$. The* determination-based correlation coefficient *(DBCC) is*

$$\eta_{\text{gen}}^2 = 1 - \frac{\sum_{k=1}^{K} P(X \in B_k) \sum_{j=1}^{c_k} \pi_{k,j} \sigma_{k,j}^2}{\text{Var}(Y)}. \tag{2}$$

5.2 Why Soft Assignment is Preferable

The soft formulation removes the four limitations of the kernel-based approach. The functional $T(\boldsymbol{\pi}, \boldsymbol{\sigma}^2) = \sum_j \pi_j \sigma_j^2$ is a smooth function of mixture parameters on the simplex times the positive orthant, and depends continuously on the data through the parameter estimates, which restores standard parametric asymptotic theory. At branch crossings the responsibilities are $\gamma_{i,k,j} \approx 1/2$ for both branches, correctly reflecting genuine uncertainty. The estimand T is symmetric under permutation of component labels, so no label-alignment post-processing is needed. Finally, setting $c_{\max}$ deliberately large and using a Dirichlet prior with small concentration parameter forces redundant components to weight zero [15]; the `BayesianGaussianMixture` class of `scikit-learn` [10] with `weight_concentration_prior_type='dirichlet_process'` realises this mechanism, so c_k need not be selected by an information criterion.

Remark 5.1. *We use Bayesian mixtures rather than maximum-likelihood EM for two reasons. First, the variational fit is well-defined when a component would otherwise collapse onto a point ($\sigma^2 \to 0$); priors regularise the variance away from zero. Second, the variational posterior over weights shrinks redundant components automatically, removing the model-selection problem. The asymptotic theory (Section 6) shows that the variational point estimates share the limiting behaviour of the maximum-likelihood estimators, keeping the proofs classical.*

5.3 Position of DBCC within the Comparison Framework

We now place DBCC in the framework of Section 2. The classical correlation ratio satisfies Rényi's identity (1) with $\mathcal{T}_\eta = \{(f, g) : f \text{ Borel}, g \equiv \text{Y}\}$. The generalised coefficient enlarges this class by allowing, besides an arbitrary transformation of X, a discrete categorisation $h : \mathbb{R} \to \{1, \dots, c\}$ of Y into branches:

$$\eta_{\text{gen}}(Y \mid X) = \sup_{f,\, h} \big|\rho\big(f(X, h(Y)),\, Y\big)\big|,$$

so that $\mathcal{T}_{\eta_{\text{gen}}} = \{(f, g) : f \text{ Borel in } (X, Z),\, g \equiv \text{Y}\}$, which strictly contains $\mathcal{T}_\eta$ (the single-branch case) and is strictly contained in the class of the maximal correlation. This yields the embedding hierarchy

$$\mathcal{T}_\rho \subsetneq \mathcal{T}_\eta \subsetneq \mathcal{T}_{\eta_{\text{gen}}} \subsetneq \mathcal{T}_{\rho^{\max}}. \tag{3}$$

By Property 7, each inclusion corresponds to a class of dependencies that the smaller coefficient underestimates: DBCC dominates the classical correlation ratio on every multivalued dependence while remaining bounded above by the maximal correlation.

Proposition 5.1. *The coefficient η_{gen}^2 satisfies the normalisation, affine-invariance, strict-dependence, and linear-consistency properties of the framework. In particular, for a bivariate normal pair with Pearson correlation ρ, one has $\eta_{\text{gen}}^2(Y \mid X) = \rho^2$.*

Proof. Normalisation. Within-component variances are non-negative and bounded above by the total variance, so the ratio in (2) lies in $[0, 1]$.

Affine invariance. Replacing Y by $a_2 Y + b_2$ scales every $\sigma_{k,j}^2$ and $\text{Var}(Y)$ by a_2^2, leaving the ratio unchanged; replacing X by $a_1 X + b_1$ relabels the bins through a monotone transformation and leaves the conditional distributions invariant.

Strict dependence. If Y is an almost-sure (possibly multivalued) function of X, the conditional variance within each branch vanishes, so the numerator is zero and $\eta^2_{\text{gen}} = 1$. Conversely $\eta^2_{\text{gen}} = 1$ forces $\mathbb{E}[\text{Var}(Y \mid X, Z)] = 0$, i.e. Y is an almost-sure function of (X, Z).

Linear consistency. For a bivariate normal pair, $Y \mid X = x$ is Gaussian with mean linear in x and constant variance $\sigma^2_Y(1-\rho^2)$. The conditional law is unimodal, so the fitted mixture collapses to one active component and $\eta^2_{\text{gen}} = 1 - \sigma^2_Y(1-\rho^2)/\sigma^2_Y = \rho^2$, reproducing Rényi's seventh axiom in directional form. □

Remark 5.2. *The independence criterion holds in the direction "independence implies $\eta^2_{\text{gen}} = 0$". The converse fails in general, since $\eta^2_{\text{gen}} = 0$ only certifies that X and the branch label do not reduce the variance of Y; this is the directional limitation already present in the classical correlation ratio. Because all coefficients in (3) share linear consistency, they are calibrated to a common scale on normal data, and their contrasts are diagnostic: $\eta^2 \approx \eta^2_{\text{gen}}$ signals a single-branch dependence, $\eta^2 \ll \eta^2_{\text{gen}}$ signals multivaluedness, $\eta^2_{\text{gen}} \ll (\rho^{\max})^2$ signals that a transformation of Y is required, and a gap between $\eta^2_{\text{gen}}(Y \mid X)$ and $\eta^2_{\text{gen}}(X \mid Y)$ reveals the direction of determination.*

6 Asymptotic Theory

This section establishes the consistency of the estimator $\widehat{\eta}^2_{\text{gen}}$ of Definition 5.1. We use an *oracle decomposition* isolating the algorithmic part of the estimator (the local mixture fits) from its statistical part (averages of identically distributed quantities), so that the difficult content is concentrated in a single lemma comparing the estimator with an idealised oracle, while the rest reduces to the law of large numbers and the continuous mapping theorem. Throughout we use *quantile-based binning*: the sample is partitioned into K_N bins each containing $\lfloor N/K_N \rfloor$ observations, the schema used in our implementation, which removes the asymmetry caused by data-sparse tails of the marginal of X.

6.1 Setup and Assumptions

Let $(X_i, Y_i)_{i=1}^N$ be an i.i.d. sample. The bins $\{B_{k,N}\}_{k=1}^{K_N}$ are defined by empirical quantiles of X, each containing $N_k = \lfloor N/K_N \rfloor$ observations. Inside each bin a finite Bayesian Gaussian mixture is fitted, yielding $(\widehat{\pi}_{k,j}, \widehat{\mu}_{k,j}, \widehat{\sigma}^2_{k,j})$. Define

$$A_N = \frac{1}{N}\sum_{k=1}^{K_N} N_k \sum_{j=1}^{c_{\max}} \widehat{\pi}_{k,j}\widehat{\sigma}^2_{k,j}, \qquad B_N = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2, \qquad \widehat{\eta}^2_{\text{gen}} = 1 - \frac{A_N}{B_N}.$$

Let $Z \in \{1, \dots, c\}$ be the latent branch indicator and write $\mu_z(x) = \mathbb{E}[Y \mid X = x, Z = z]$. The estimand is

$$\eta^2_{\text{gen}} = 1 - \frac{\mathbb{E}[\text{Var}(Y \mid X, Z)]}{\text{Var}(Y)}. \tag{4}$$

Because the latent branch Z is a genuine variable of the generative model, (4) is the variance of Y explained by X and Z *jointly*; Section 6.5 shows how to split this into the part intrinsic to the marginal law of Y and the part attributable to X, which restores a clean independence criterion without any restriction on the distribution of Y. We adopt the following assumptions.

(A1) *Moments.* (X_i, Y_i) are i.i.d. with $\mathbb{E}[Y^4] < \infty$ and $\text{Var}(Y) > 0$.

(A2) *Bin schedule.* The bins are quantile-based with $N_k = \lfloor N/K_N \rfloor$, $K_N \to \infty$ and $K_N/N \to 0$. The cube-root rule $K_N \asymp N^{1/3}$ is the default; the practical choice is discussed in Section 8.

(A3) *Identifiability.* The Gaussian component family is identifiable in the sense of Teicher [20]: two finite Gaussian mixtures with the same density agree up to a permutation of labels.

(A4) *Component family.* For every x the conditional density $p(y \mid x)$ admits a finite-Gaussian-mixture representation; in the misspecified case the estimand is replaced by its Kullback–Leibler projection onto the mixture family.

(A5) *Variational consistency of the functional.* For a fixed bin with $n \to \infty$, the variational estimate of the residual-variance functional $T(\boldsymbol{\pi}, \boldsymbol{\sigma}^2) = \sum_j \pi_j \sigma_j^2$ is consistent, $\widehat{T}_n^{\mathrm{VB}} \xrightarrow{p} T^*$. We state the assumption at the level of T rather than of the full parameter vector because we work in the overfitted regime $c_{\max} > c$, where the component-wise mixture asymptotics are non-regular: redundant components carry weights tending to zero [15], so the individual parameters need not admit a regular expansion. The functional T is nevertheless continuous on the closure of the parameter space and invariant to the vanishing components, so its consistency follows from the frequentist consistency of variational Bayes [23] applied to the active components.

Remark 6.1. *Quantile binning in (A2) guarantees $N_k \to \infty$ uniformly across bins by construction, removing any need for a regularity condition on the support or density of X. Because each N_k is equal, all aggregation steps reduce to full-sample averages, for which the moment condition in (A1) suffices; no separate uniform-integrability assumption is required.*

Remark 6.2. *Stating (A5) for the smooth, permutation-symmetric functional T rather than for the raw parameters is deliberate. In the overfitted regime the individual component locations and weights are not identifiable in a regular way, so a first-order expansion of $\widehat{\theta}^{\mathrm{VB}}$ in the manner of the regular case would overreach the cited results. The quantity actually evaluated, $T = \sum_j \pi_j \sigma_j^2$, depends only on the active mixture and is continuous as redundant weights vanish; this is all that Lemma 6.2 and Theorem 6.1 require. The stronger parameter-level expansion is invoked only inside the conditional assumption (A6) for asymptotic normality.*

6.2 The Oracle Estimator

Definition 6.1. *The* oracle estimator *of the population residual variance is*

$$\widetilde{A}_N = \frac{1}{N} \sum_{i=1}^{N} \big(Y_i - \mu_{Z_i}(X_i)\big)^2. \tag{5}$$

The oracle is not computable: it depends on the latent Z_i and the population functions μ_z. It serves as an intermediate target, close to $\mathbb{E}[\mathrm{Var}(Y \mid X, Z)]$ by an elementary argument, with the computable A_N shown close to it by a single lemma.

Lemma 6.1. *Under* (A1), $\widetilde{A}_N \xrightarrow{p} \mathbb{E}[\mathrm{Var}(Y \mid X, Z)]$.

Proof. The summands $W_i = (Y_i - \mu_{Z_i}(X_i))^2$ are i.i.d. with $\mathbb{E}[W_i] = \mathbb{E}[\mathrm{Var}(Y \mid X, Z)]$, since $\mathbb{E}[Y \mid X, Z] = \mu_Z(X)$. The fourth-moment condition in (A1) gives $\mathbb{E}[W_i^2] < \infty$, so the weak law of large numbers applies. □

6.3 The BMM–Oracle Gap

Lemma 6.2. *Under* (A1)–(A5) *with the cube-root rule $K_N \asymp N^{1/3}$, $A_N - \widetilde{A}_N \xrightarrow{p} 0$.*

Proof. Let $k(i)$ index the bin containing X_i, and let $T_k^* = \sum_j \pi_{k,j}^* \sigma_{k,j}^{*\,2}$ be the population mixture functional in bin k, with $\widehat{T}_k$ its variational estimate. With $\bar{A}_N = N^{-1} \sum_k N_k T_k^*$ we split

$$A_N - \widetilde{A}_N = \underbrace{(A_N - \bar{A}_N)}_{\text{estimation error}} + \underbrace{(\bar{A}_N - \widetilde{A}_N)}_{\text{discretisation error}} .$$

Estimation error. By (A2) each bin has $N_k \asymp N^{2/3} \to \infty$. For fixed k, the population functional T_k^* is identified under (A3) and (A4), and the functional $T(\boldsymbol{\pi}, \boldsymbol{\sigma}^2) = \sum_j \pi_j \sigma_j^2$ is continuous and permutation-symmetric, so (A5) gives $\widehat{T}_k \xrightarrow{p} T_k^*$ for each bin. Since $N_k/N = 1/K_N$,

$$|A_N - \bar{A}_N| \le \max_{1 \le k \le K_N} |\widehat{T}_k - T_k^*| = O_p\big(\sqrt{\log N / N^{2/3}}\big) = o_p(1),$$

by the uniform consistency of the functional estimate over sub-samples of size N_k [1, 12, 22].

Discretisation error. Here $T_k^* = \mathbb{E}[\mathrm{Var}(Y \mid X \in B_{k,N}, Z)]$ and $N_k/N = 1/K_N = \widehat{P}(X \in B_{k,N})$, so $\bar{A}_N$ is the empirical-measure approximation to $\mathbb{E}[\mathrm{Var}(Y \mid X, Z)]$. Writing

$$\bar{A}_N - \widetilde{A}_N = \sum_{k=1}^{K_N} \frac{1}{K_N} \Big(T_k^* - \frac{1}{N_k} \sum_{i:k(i)=k} (Y_i - \mu_{Z_i}(X_i))^2 \Big),$$

each bracket is $O_p(N_k^{-1/2}) = O_p(N^{-1/3})$ by the L^2 weak law, so the weighted sum is $O_p(N^{-1/3}) = o_p(1)$. Combining the two terms gives the claim. □

6.4 Consistency

Theorem 6.1. *Under* (A1)–(A5),

$$\widehat{\eta}^2_{\mathrm{gen}} \xrightarrow{p} \eta^2_{\mathrm{gen}} = 1 - \frac{\mathbb{E}[\mathrm{Var}(Y \mid X, Z)]}{\mathrm{Var}(Y)} \qquad (N \to \infty).$$

Proof. By Lemmas 6.1 and 6.2 and the triangle inequality, $A_N \xrightarrow{p} \mathbb{E}[\mathrm{Var}(Y \mid X, Z)]$. By (A1) and the weak law, $B_N \xrightarrow{p} \mathrm{Var}(Y) > 0$. The map $g(a, b) = 1 - a/b$ is continuous on $\{b > 0\}$, so the continuous mapping theorem [21] gives $\widehat{\eta}^2_{\mathrm{gen}} = g(A_N, B_N) \xrightarrow{p} g(A, B) = \eta^2_{\mathrm{gen}}$. □

Remark 6.3. *The bound in Lemma 6.2 is conservative: it controls the worst bin and yields the crude rate* $|\widehat{\eta}^2_{\mathrm{gen}} - \eta^2_{\mathrm{gen}}| = O_p(N^{-1/3})$. *Averaging over the independent bin-level fits in fact recovers the parametric* $N^{-1/2}$ *rate up to a discretisation bias of order* K_N^{-2}, *which is what underlies the asymptotic normality of Section 6.6.*

6.5 Attributing the Explained Variance to X and Z

The estimand (4) conditions on both X and the latent branch Z, so it reports the variance explained by the two jointly. This is the honest content of the coefficient - Z is a real latent variable, not an artefact - but it conflates two distinct sources: the structure intrinsic to the marginal law of Y (the mere existence of branches) and the extent to which X actually *locates* those branches. The law of total variance separates them exactly, without restricting the distribution of Y.

Applying the law of total variance to Y given (X, Z) and then to Y given Z alone,

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X, Z)] + \mathrm{Var}(\mu_Z(X)), \qquad \mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid Z)] + \mathrm{Var}(\mathbb{E}[Y \mid Z]).$$

Write the joint and the marginal-branch coefficients as

$$\eta^2_{(X,Z)} = \frac{\mathrm{Var}(\mu_Z(X))}{\mathrm{Var}(Y)} \big(= \eta^2_{\mathrm{gen}}\big), \qquad \eta^2_Z = \frac{\mathrm{Var}(\mathbb{E}[Y \mid Z])}{\mathrm{Var}(Y)}.$$

Since $\mathbb{E}[\mu_Z(X) \mid Z] = \mathbb{E}[Y \mid Z]$, a third application of the law of total variance to $\mu_Z(X)$ conditional on Z yields the exact orthogonal split

$$\eta^2_{(X,Z)} = \eta^2_Z + \eta^2_{X|Z}, \qquad \eta^2_{X|Z} = \frac{\mathbb{E}\big[\mathrm{Var}(\mu_Z(X) \mid Z)\big]}{\mathrm{Var}(Y)} \geq 0. \tag{6}$$

The term η^2_Z is the variance explained by the branch label alone - the intrinsic multimodality of Y - while $\eta^2_{X|Z}$ is the variance explained by the movement of the branch centres $\mu_Z(X)$ as X varies within a branch: precisely the contribution of X. For a closed curve such as the circle the centres sweep with X and $\eta^2_{X|Z}$ is large; for an X-independent bimodal Y the centres are fixed and $\eta^2_{X|Z} = 0$, even though $\eta^2_{(X,Z)}$ is large.

The identity (6) is a population statement that mentions the latent Z. Whether its two components can be *estimated* separately depends on what is observed about Z, and we distinguish two regimes.

Regime 1: the branch label is observed. If a label Z_i is available for each observation - as in a designed experiment, or when the branches correspond to a known categorical covariate - the decomposition is computed directly, with neither a mixture nor a permutation. One estimates the marginal term by conditioning on the labels,

$$\widehat{\eta}^2_Z = 1 - \frac{\sum_z \widehat{P}(Z = z)\, \widehat{\mathrm{Var}}(Y \mid Z = z)}{\widehat{\mathrm{Var}}(Y)}, \tag{7}$$

and the joint term by the oracle estimator (5), which is now feasible because the Z_i are known; their difference estimates $\eta^2_{X|Z}$. In this regime DBCC reduces to a two-way (between-branch and within-branch) variance decomposition, and the mixture machinery of Section 5 is unnecessary. Both estimators are consistent by the law of large numbers under (A1). This case is clean but uncommon: the labels are rarely available.

Regime 2: the branch label is latent. In most applications Z is unobserved - recovering it is the very reason for the mixture. The joint coefficient $\eta^2_{(X,Z)}$ is then estimated by $\widehat{\eta}^2_{\mathrm{gen}}$ as before, the mixture supplying soft surrogates for the unknown Z. The marginal term η^2_Z cannot be read off directly, since the mixture is fitted within bins of X and so already mixes the two contributions. A random permutation of the predictor recovers it without any knowledge of Z.

Definition 6.2. *Let π be a uniform random permutation of the predictor values, which breaks the association between X and Y while preserving both marginals. The* permutation-corrected *coefficient is*

$$\Delta\eta^2_{\mathrm{gen}}(Y \mid X) = \eta^2_{\mathrm{gen}}(Y \mid X) - \mathbb{E}_\pi\, \eta^2_{\mathrm{gen}}\big(Y \mid \pi(X)\big). \tag{8}$$

Proposition 6.1. *Under* (A1)–(A5), $\mathbb{E}_\pi\, \eta^2_{\mathrm{gen}}(Y \mid \pi(X)) \xrightarrow{p} \eta^2_Z$ *and hence* $\Delta\eta^2_{\mathrm{gen}}(Y \mid X) \xrightarrow{p} \eta^2_{X|Z}$. *In particular, if $X \perp Y$ then $\eta^2_{X|Z} = 0$ and $\Delta\eta^2_{\mathrm{gen}}(Y \mid X) \xrightarrow{p} 0$, for any marginal law of Y with finite fourth moment.*

Proof. The permuted sample is distributed as an independent coupling $X' \perp Y$ with the same marginals, in which each bin contains an asymptotically i.i.d. draw from the marginal of Y. The

within-bin mixture therefore fits the marginal law, and by Theorem 6.1 applied to the permuted sample $\eta^2_{\text{gen}}(Y \mid \pi(X)) \xrightarrow{p} 1 - \mathbb{E}[\text{Var}(Y \mid Z)]/\,\text{Var}(Y) = \eta^2_Z$. Subtracting from $\eta^2_{\text{gen}}(Y \mid X) \xrightarrow{p} \eta^2_{(X,Z)}$ and using (6) gives the limit $\eta^2_{X|Z}$. If $X \perp Y$ then $X \perp (Y, Z)$, because Z is a function of Y; thus $\mu_Z(X) = \mathbb{E}[Y \mid Z]$ is constant in X within each branch, so $\text{Var}(\mu_Z(X) \mid Z) = 0$ and $\eta^2_{X|Z} = 0$. □

Remark 6.4. *The permutation route estimates the same marginal term η^2_Z that the labelled estimator (7) computes when Z is known, but does so Z-free. A simpler alternative in Regime 2 would fit a single mixture to the pooled Y and read off η^2_Z directly; the permutation is preferable because it reuses the exact binned pipeline of $\widehat{\eta}^2_{\text{gen}}$, so the finite-sample biases of binning and of the mixture fit largely cancel in the difference (8). In practice the expectation is estimated by a small number B of permutations ($B = 5$ to 20 suffices), each costing one DBCC evaluation, so the correction preserves linear time.*

Remark 6.5. *The decomposition is empirically tight. On a noisy circle the joint and marginal terms are about 0.99 and 0.81, giving $\Delta\eta^2_{\text{gen}} \approx 0.19$, whereas on an X-independent bimodal Y both terms equal about 0.90 and $\Delta\eta^2_{\text{gen}} \approx 0$. The raw coefficient η^2_{gen} remains the quantity of interest when the branch structure is known to be induced by X, as for the multivalued relations of Section 9; the corrected $\Delta\eta^2_{\text{gen}}$ is the conservative default whenever the intrinsic multimodality of Y cannot be ruled out - the common situation, since Z is latent - and it restores the independence criterion in full generality.*

6.6 Asymptotic Normality

Consistency does not by itself license confidence intervals; for that we need the limiting distribution. The oracle representation supplies the leading influence function, and one further regularity condition controls the bin structure.

(A6) *Asymptotic normality of the base statistics.* There is a positive semi-definite matrix $\Sigma = (\Sigma_{ab})_{a,b\in\{1,2\}}$ such that

$$\sqrt{N}\begin{pmatrix} A_N - A \\ B_N - B \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma), \qquad A = \mathbb{E}[\text{Var}(Y \mid X, Z)],\ B = \text{Var}(Y).$$

Remark 6.6. *We stress that (A6)* assumes *the hardest part of the problem - the joint $\sqrt{N}$-asymptotic normality of the base statistics (A_N, B_N) for a partitioning estimator with a growing number of cells, including the contribution of the within-bin variational fits to the limiting covariance. We do not prove (A6) here, so Theorem 6.2 below is* a conditional *result: a delta-method corollary of (A6), not a self-contained proof of normality. Its value is to identify the limiting form and the asymptotic variance once (A6) is granted, and to justify the bootstrap. As for when (A6) holds: under (A2), if the partition undersmooths enough that the discretisation bias is negligible at the $\sqrt{N}$ scale - when $x \mapsto \mathbb{E}[\text{Var}(Y \mid X = x, Z)]$ is Lipschitz the bias is $O(K_N^{-2})$, which is $o(N^{-1/2})$ for $K_N \asymp N^{\alpha}$ with $\alpha > 1/4$, so the cube-root rule qualifies - the leading term of A_N is the oracle average $N^{-1}\sum_i (Y_i - \mu_{Z_i}(X_i))^2$ and that of B_N is $N^{-1}\sum_i (Y_i - \bar{Y})^2$, so Σ is governed by the joint law of $W = (Y - \mu_Z(X))^2$ and $V = (Y - \mathbb{E}Y)^2$, with an additional contribution from the within-bin estimation.*

Theorem 6.2. *Under* (A1)–(A6) *there exists $\sigma^2_\eta \geq 0$ with*

$$\sqrt{N}\,(\widehat{\eta}^2_{\text{gen}} - \eta^2_{\text{gen}}) \xrightarrow{d} \mathcal{N}(0, \sigma^2_\eta), \qquad \sigma^2_\eta = \frac{\Sigma_{11}}{B^2} - \frac{2A\,\Sigma_{12}}{B^3} + \frac{A^2\,\Sigma_{22}}{B^4}.$$

Proof. The map $g(a,b) = 1 - a/b$ is continuously differentiable on $\{b > 0\}$, with gradient $\nabla g(A,B) = (-1/B,\ A/B^2)^\top$. Since $B > 0$ by (A1), the delta method applied to (A6) gives $\sqrt{N}(\widehat{\eta}^2_{\text{gen}} - \eta^2_{\text{gen}}) \xrightarrow{d} \mathcal{N}(0, \nabla g^\top \Sigma \nabla g)$, and expanding the quadratic form yields the stated σ^2_η. □

Remark 6.7. *The entries of Σ involve fourth-order functionals of the branch residuals and the within-bin estimation variance, which are awkward to evaluate analytically. For confidence intervals and significance tests we therefore recommend the nonparametric bootstrap [2], whose consistency for smooth functionals of M-estimators applies here;* 200 *to* 500 *resamples typically suffice.*

7 Robustness

Like every variance-based measure, DBCC is sensitive to extreme outliers. When a single observation has $|Y_i| \to \infty$, both A_N and B_N grow, the ratio $A_N/B_N \to 1$, and the coefficient is driven to zero. The influence function is unbounded and the asymptotic breakdown point is 0%. This is a structural property of any L_2 measure, shared by Pearson correlation and the classical correlation ratio.

Remark 7.1. *A natural robust variant replaces variance by the squared median absolute deviation* $\text{MAD}^2(Y) = (\text{median}\,|Y_i - \text{median}(Y)|)^2$. *Squared MAD preserves quadratic scale, so the variant is numerically comparable to η^2_{gen} on clean data, since $\sigma \approx 1.4826\,\text{MAD}$ for normal samples and the constant cancels in the ratio. The bounded influence function of MAD raises the breakdown point to* 50%. *A companion modification within the mixture is to use Student-t components, whose location and scale are themselves robust. A detailed analysis is left for future work.*

8 Computational Complexity and Practical Aspects

Proposition 8.1. *Under* (A2) *with $K_N \asymp N^{1/3}$, and with the number of components $c_{\max}$ and variational iterations T treated as constants, the soft-assignment DBCC has total computational complexity $\mathcal{O}(N)$.*

Proof. The total cost is the sum over bins of the cost of one mixture fit. Variational fitting of a mixture with $c_{\max}$ components on N_k observations costs $\mathcal{O}(c_{\max} N_k T)$. With $K_N \asymp N^{1/3}$ and $N_k \asymp N^{2/3}$,

$$\text{cost} = \mathcal{O}(N^{1/3} \cdot c_{\max} \cdot N^{2/3} \cdot T) = \mathcal{O}(c_{\max} T N),$$

which is $\mathcal{O}(N)$ for constant $c_{\max}, T$. The same holds for any schedule $K_N \asymp N^\alpha$, $0 < \alpha < 1$, since $K_N \cdot N_k = N$ identically. The permutation correction of (8) multiplies the cost by the constant $B + 1$ and so preserves linear time. □

The linear-time character distinguishes DBCC from $\mathcal{O}(N^2)$ measures such as distance correlation [19] and from the exact MIC [14], which is super-polynomial in its grid-resolution parameter.

Remark 8.1. *The cube-root rule $K_N = \lceil N^{1/3} \rceil$ is the default: it satisfies the requirements of Theorem 6.1 and matches the mean-integrated-square-error-optimal histogram bin width [17]. Quantile-based binning, with equal counts per bin, is preferable in heavy-tailed regimes, improving local sample sizes in sparse tails. The overfitted-prior mechanism makes $c_{\max}$ unimportant once it is not too small; we recommend $c_{\max} = 5$, with the small default Dirichlet concentration of* `scikit-learn` *[10] and several random restarts for stability on small bins.*

Remark 8.2 (Undersmoothing for fast-changing relations)**.** *The cube-root rule is tuned for density estimation, not for resolving the conditional mean within a bin. When* $\mathbb{E}[Y \mid X = x]$ *varies strongly across a bin - as for the exponential relation, or whenever* N *is moderate so that a coarse partition results - the within-bin variation of the mean is absorbed into the local variance and inflates it, which destabilises the mixture fit and makes the coefficient erratic. The remedy is a finer partition. At* $N = 10^4$*, for example, the cube-root rule yields only about* 21 *to* 22 *bins of some* 450 *to* 500 *points each, whereas the square-root rule* $K_N = \lceil N^{1/2} \rceil$ *yields* 100 *bins of about* 100 *points each; the finer partition makes the conditional mean nearly constant within each bin, so the local variance reflects the branch structure rather than the trend. This is a deliberate* undersmoothing*: it suppresses the within-bin trend bias at the cost of fewer points per bin and a slightly slower asymptotic rate. We therefore recommend the square-root rule for relations with strong local curvature and for small to moderate samples, and the cube-root rule when the conditional mean is gentle and* N *is large. Both choices preserve the linear-time complexity of Proposition 8.1 and the consistency of Theorem 6.1, which requires only* $K_N \to \infty$ *and* $K_N/N \to 0$*.*

9 Experiments

This section reports a synthetic study with two purposes: to describe how the experiment was conducted, and to interpret the results as evidence that the proposed coefficient handles multivalued dependencies - such as the double sinusoid, the circle, and circular clusters - on which the competing measures are less effective or fail entirely.

9.1 Experimental Design

Each experiment fixes a dependence type and a noise level and proceeds in three steps: data generation, evaluation of all measures, and repetition across an increasing grid of noise levels.

Data generation. For every dependence type we generate $N = 10\,000$ independent observations (X_i, Y_i). The predictor X is drawn from a fixed marginal distribution and the response is produced by the corresponding generating mechanism with additive Gaussian noise of standard deviation σ_n. The noise level is swept over the range $\sigma_n \in [0.01, 6]$, and the data are regenerated independently for each value of σ_n, so that the reported curves reflect the behaviour of the measures rather than the peculiarities of a single random sample.

Dependence types. The study spans twelve qualitatively distinct relations. Six are single-valued: *linear* $Y = aX + \varepsilon$ (baseline); *quadratic* $Y = X^2 + \varepsilon$ (symmetric nonlinear, on which Pearson vanishes); *exponential* and *logarithmic* (monotonic nonlinear); *sigmoid* (saturating monotonic); and *sinusoidal* $Y = \sin(\omega X) + \varepsilon$ (non-monotonic but single-valued). Six are threshold or multivalued: *step* and *hill* (sharp transitions); *double sinusoidal* (two overlapping branches with periodic crossings); *circle* (a closed curve, the canonical multivalued case); *circle clusters* (a circle with additional hidden categorical structure); and the *null* case $X \perp Y$ (a baseline that measures the value each coefficient attains on genuinely independent data).

Measures compared. On each generated dataset we evaluate six quantities: the proposed determination-based coefficient η^2_{gen}; the classical correlation ratio η^2, which serves as the single-branch reference; the Randomized Dependence Coefficient (RDC) [7] using the implementation of Doran [4]; the distance correlation (dCor) [19] using the `dcor` package [11]; the Pearson correlation r [9]; and the Spearman rank correlation ρ_S [18]. The proposed coefficient is computed as in Section 5, with the square-root quantile-binning rule and a Bayesian Gaussian mixture ($c_{\max} = 5$, Dirichlet-process prior) from `scikit-learn` [10]. Pearson and Spearman are reported as signed values in $[-1, 1]$; the remaining measures lie in $[0, 1]$.

For every dependence type we plot all six measures as functions of the noise level σ_n on a shared vertical scale, so that the curves can be read against a common baseline. As σ_n increases each

coefficient degrades from its noise-free value towards the level it attains on independent data, and the relative ordering of the curves quantifies the differential sensitivity of the measures to additive noise.

9.2 Interpretation of the Results

Figure 1 collects the six single-valued relations and Figure 2 the six threshold and multivalued relations.

Single-valued relations. On the linear relation (Figure 1a) all measures agree at low noise and decline together as σ_n grows, confirming that the proposed coefficient is calibrated to the same scale as the classical measures on data within their functional class. On the quadratic relation (Figure 1b) Pearson and Spearman collapse to zero - the symmetric parabola produces no linear or monotone trend - while η^2_{gen}, η^2, dCor, and RDC all remain high, illustrating the suboptimality of the linear and rank measures outside their functional class. The monotonic nonlinear relations (exponential, logarithmic, sigmoid) and the oscillatory sinusoid (Figures 1c–1f) are single-valued, so the classical correlation ratio η^2 already captures them; here the proposed coefficient tracks η^2 closely, as expected, since for a single-branch dependence the two coincide.

Threshold and multivalued relations. The threshold relations, step and hill (Figures 2a and 2b), are single-valued and are handled comparably by η^2_{gen} and η^2, the proposed coefficient remaining slightly more stable as the noise grows. The decisive comparison is provided by the genuinely multivalued relations. On the double sinusoid (Figure 2c), the circle (Figure 2d), and the circular clusters (Figure 2e), a single value of X corresponds to several values of Y; the conditional mean is therefore nearly constant, and every single-branch measure - the classical correlation ratio η^2, Pearson, Spearman, dCor, and RDC - sits near zero across the whole noise range. The proposed coefficient η^2_{gen}, by contrast, starts close to unity and remains clearly separated from the others until the noise level is large enough to merge the branches, at which point it too decays to the baseline. This separation is the central empirical finding of the paper: the soft-assignment mixture decomposition recovers the residual variance within each branch, so the coefficient registers the strong dependence that the conditional mean alone cannot see.

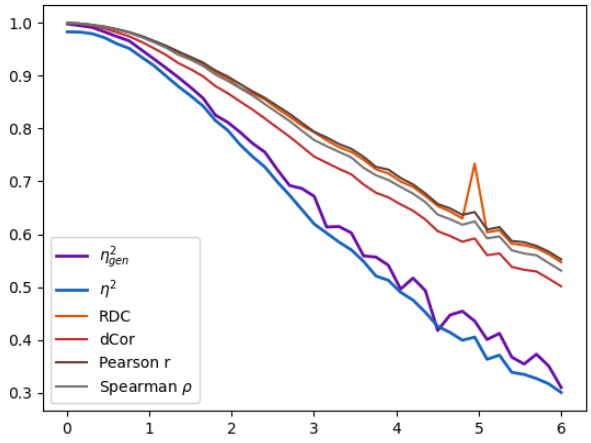
Null case. On independent data (Figure 2f) all six measures fluctuate around zero with magnitude controlled only by sampling error. In particular the proposed coefficient does not manufacture spurious dependence on the null relation, which confirms that its high values on the multivalued relations reflect genuine structure rather than an artefact of the mixture fit.

The DBCC implementation, the synthetic data generator, and the plotting scripts that produced Figures 1 and 2 are available in the supplementary material.

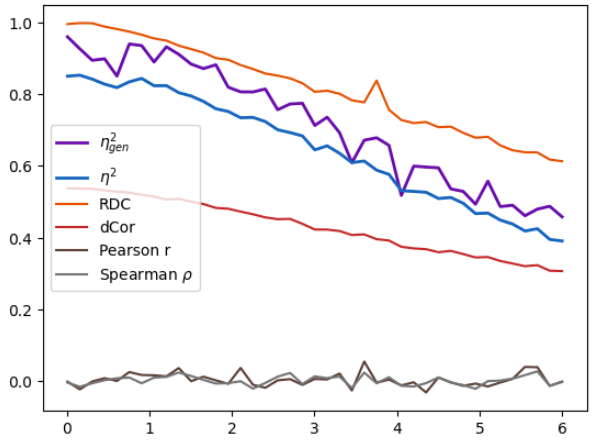
10 Conclusions and Future Work

This paper developed a determination-based correlation coefficient (DBCC) for multivalued dependencies, together with a unified framework for comparing dependence measures of different mathematical natures.

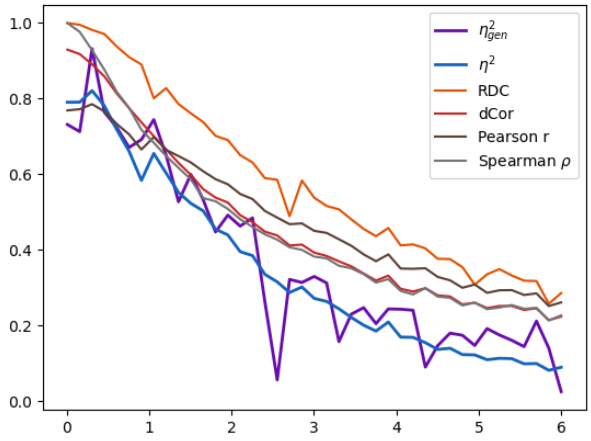
The framework rests on the functional class of a coefficient and seven linking properties anchored to the Pearson correlation, providing macro-level common ground on which otherwise incommensurable coefficients can be compared, and identifying the suboptimality criterion as the source of systematic underestimation. The DBCC itself replaces the discontinuous kernel-based mode-seeking of an earlier version with a soft-assignment Bayesian Gaussian mixture model, retaining the variance-explained interpretation of R^2 while extending it to closed curves, ellipses, and hidden categorical structure. We proved consistency via an oracle decomposition that isolates the algorithmic from the statistical error, relying only on standard inputs and an explicit, literature-backed variational-to-MLE assumption. We established linear-time, embarrassingly



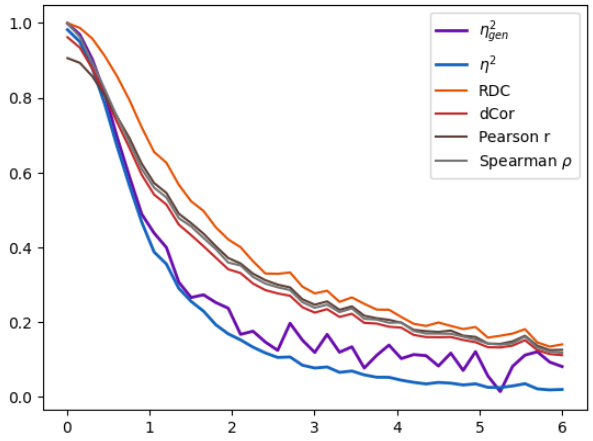
(a) Linear



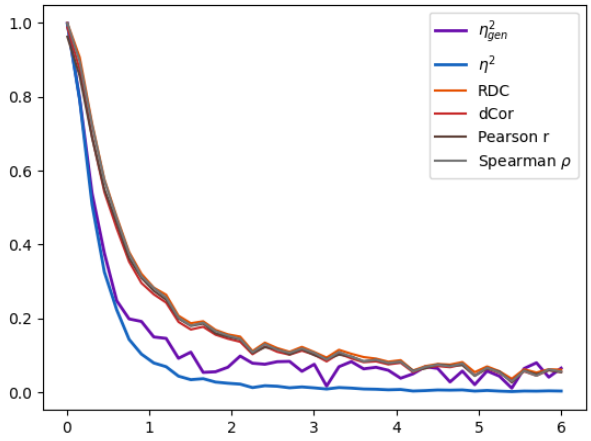
(b) Quadratic



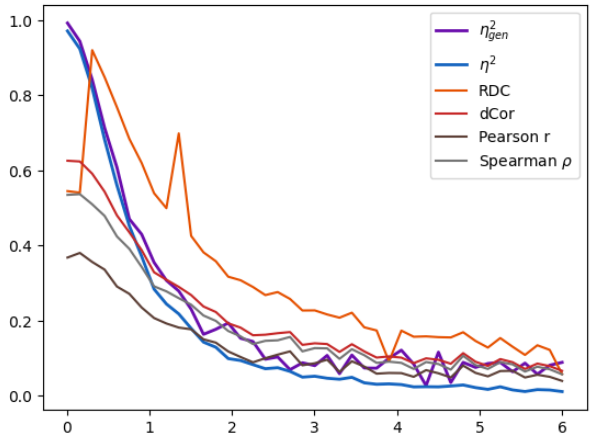
(c) Exponential



(d) Logarithmic

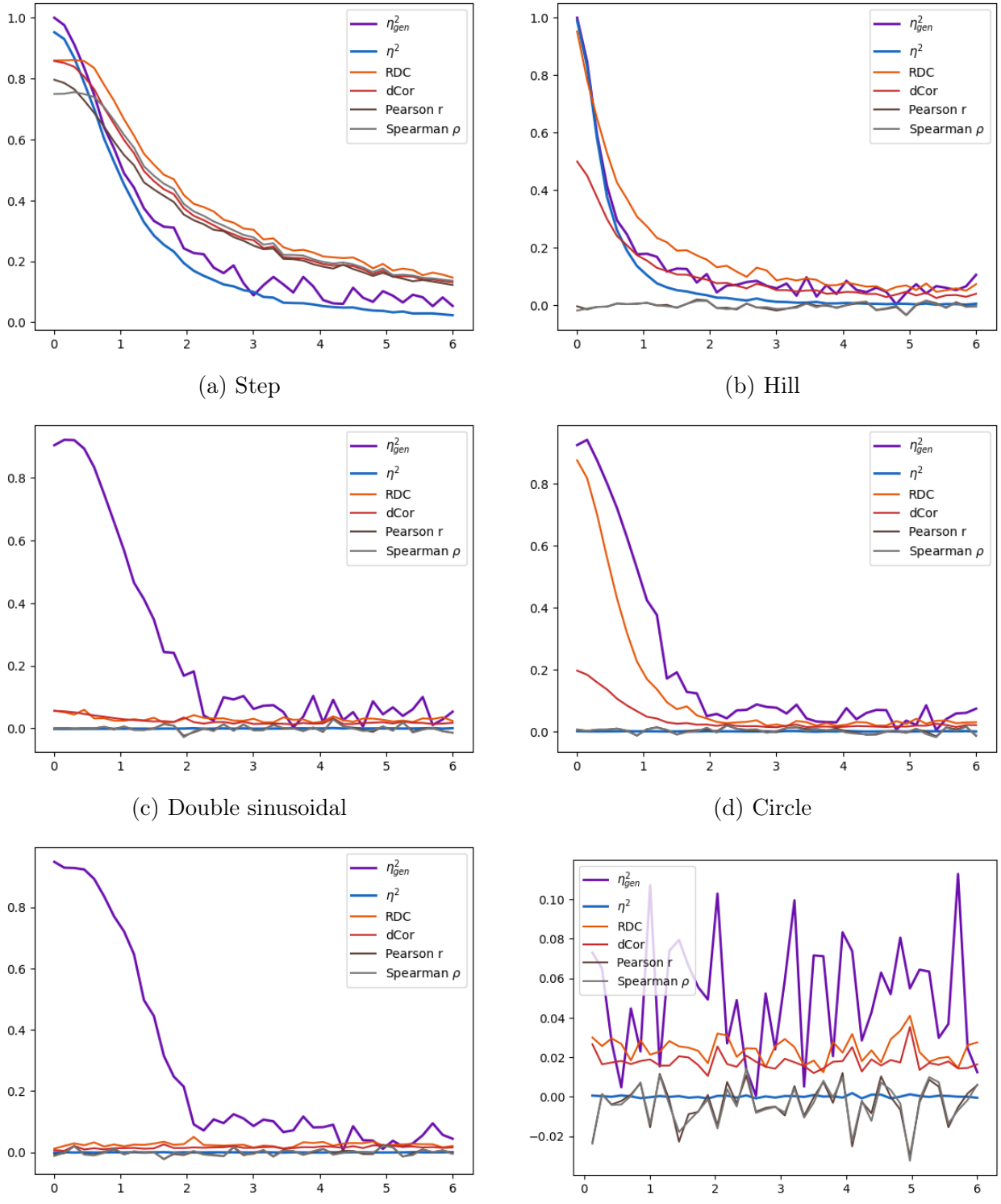


(e) Sigmoid



(f) Sinusoidal

Figure 1: Dependence measures as functions of the noise level σ_n for the single-valued relations: linear, quadratic, exponential, logarithmic, sigmoid, and sinusoidal. Each panel overlays the proposed coefficient η^2_{gen} and the comparison measures (η^2, RDC, dCor, Pearson r, Spearman ρ) on a common vertical scale.



(a) Step (b) Hill

(c) Double sinusoidal (d) Circle

(e) Circle clusters (f) None (independent)

Figure 2: Dependence measures as functions of the noise level σ_n for the threshold and multivalued relations: step, hill, double sinusoidal, circle, circle clusters, and the null (independent) case. On the multivalued relations (c–e) the determination-based coefficient η^2_{gen} remains high where the single-branch measures, including the classical correlation ratio η^2, collapse to near zero.

parallel computation, and positioned DBCC within the framework as a supremum of Pearson correlations over a functional class strictly between those of the correlation ratio and the maximal correlation, making its contrasts with neighbouring measures diagnostic.

Список літератури

[1] Bickel P.J., Doksum K.A. Mathematical statistics: basic ideas and selected topics, Volume I, 2nd ed. Chapman and Hall/CRC, Boca Raton, 2015. doi:10.1201/b18312

[2] Bickel P.J., Freedman D.A. *Some asymptotic theory for the bootstrap*. Ann. Statist. 1981, **9** (6), 1196–1217. doi:10.1214/aos/1176345637

[3] Chatterjee S. *A new coefficient of correlation*. J. Amer. Statist. Assoc. 2021, **116** (536), 2009–2022. doi:10.1080/01621459.2020.1758115

[4] Doran G. RDC: the randomized dependence coefficient in Python. Source code (software), 2014. https://github.com/garydoranjr/rdc (accessed June 2026).

[5] Gebelein H. *Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung*. Z. Angew. Math. Mech. 1941, **21** (6), 364–379. doi:10.1002/zamm.19410210604 (in German)

[6] Hirschfeld H.O. *A connection between correlation and contingency*. Math. Proc. Cambridge Philos. Soc. 1935, **31** (4), 520–524. doi:10.1017/S0305004100013517

[7] Lopez-Paz D., Hennig P., Schölkopf B. The randomized dependence coefficient. In: Burges C.J.C., Bottou L., Welling M., Ghahramani Z., Weinberger K.Q. (Eds.) Advances in Neural Information Processing Systems 26 (NIPS 2013). Curran Associates, 2013, 1–9.

[8] McLachlan G.J., Peel D. Finite mixture models. In: Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 2000. doi:10.1002/0471721182

[9] Pearson K. *Note on regression and inheritance in the case of two parents*. Proc. Roy. Soc. London 1895, **58**, 240–242. doi:10.1098/rspl.1895.0041

[10] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. *Scikit-learn: machine learning in Python*. J. Mach. Learn. Res. 2011, **12**, 2825–2830.

[11] Ramos-Carreño C., Torrecilla J.L. *dcor: distance correlation and energy statistics in Python*. SoftwareX 2023, **22**, 101326. doi:10.1016/j.softx.2023.101326

[12] Redner R.A., Walker H.F. *Mixture densities, maximum likelihood and the EM algorithm*. SIAM Rev. 1984, **26** (2), 195–239. doi:10.1137/1026034

[13] Rényi A. *On measures of dependence*. Acta Math. Acad. Sci. Hungar. 1959, **10** (3–4), 441–451. doi:10.1007/BF02024507

[14] Reshef D.N., Reshef Y.A., Finucane H.K., Grossman S.R., McVean G., Turnbaugh P.J., Lander E.S., Mitzenmacher M., Sabeti P.C. *Detecting novel associations in large data sets*. Science 2011, **334** (6062), 1518–1524. doi:10.1126/science.1205438

[15] Rousseau J., Mengersen K. *Asymptotic behaviour of the posterior distribution in overfitted mixture models*. J. R. Stat. Soc. Ser. B Stat. Methodol. 2011, **73** (5), 689–710. doi:10.1111/j.1467-9868.2011.00781.x

[16] Scarsini M. *On measures of concordance*. Stochastica 1984, **8** (3), 201–218.

[17] Scott D.W. *On optimal and data-based histograms*. Biometrika 1979, **66** (3), 605–610. doi:10.1093/biomet/66.3.605

[18] Spearman C. *The proof and measurement of association between two things*. Amer. J. Psychol. 1904, **15** (1), 72–101. doi:10.2307/1412159

[19] Székely G.J., Rizzo M.L., Bakirov N.K. *Measuring and testing dependence by correlation of distances*. Ann. Statist. 2007, **35** (6), 2769–2794. doi:10.1214/009053607000000505

[20] Teicher H. *Identifiability of finite mixtures*. Ann. Math. Statist. 1963, **34** (4), 1265–1269. doi:10.1214/aoms/1177703862

[21] van der Vaart A.W. Asymptotic statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, 3. Cambridge University Press, Cambridge, 1998. doi:10.1017/CBO9780511802256

[22] van der Vaart A.W., Wellner J.A. Weak convergence and empirical processes. In: Springer Series in Statistics. Springer, New York, 1996. doi:10.1007/978-1-4757-2545-2

[23] Wang Y., Blei D.M. *Frequentist consistency of variational Bayes*. J. Amer. Statist. Assoc. 2019, **114** (527), 1147–1161. doi:10.1080/01621459.2018.1473776

Received 20.06.2026

Accepted 02.11.2026

Published 02.11.2026

Голдованський В.А., Алєксєєв В.І. *Коефіцієнт кореляції на основі детермінації для багатозначних залежностей із застосуванням баєсових моделей сумішей* // Буковинський матем. журнал — 2026. — Т.14, №2. — С. 37–54.

Вимірювання статистичної залежності між двома випадковими величинами є фундаментальною задачею аналізу даних, проте жодна окрема міра не охоплює всіх типів зв'язку, залишаючись водночас інтерпретованою. Класичні коефіцієнти виявляють лише вузькі класи залежностей: коефіцієнт Пірсона — лінійні, коефіцієнт Спірмена — монотонні. Загальніші міри або обчислювально затратні, або схильні до завищення сили зв'язку, або не дають інформації про його структуру. У роботі отримано два основні результати. По-перше, запропоновано уніфіковану схему порівняння мір залежності різної математичної природи, яка ґрунтується на функціональному класі коефіцієнта та наборі зв'язувальних властивостей, прив'язаних до кореляції Пірсона як до спільного еталона. По-друге, на базі кореляційного відношення побудовано коефіцієнт кореляції на основі детермінації (DBCC), який охоплює багатозначні залежності, коли одному значенню предиктора може відповідати кілька станів відгуку. Коефіцієнт розкладає умовний розподіл у локальних інтервалах за допомогою баєсової гауссової моделі суміші з м'яким віднесенням спостережень до компонент, що замінює розривний пошук мод попередньої реалізації на основі ядерних оцінок густини. Доведено спроможність оцінки коефіцієнта за допомогою оракульного розкладу, який відокремлює алгоритмічну складову похибки від статистичної, і показано, що коефіцієнт обчислюється за лінійний час. Побудова зберігає прозору інтерпретацію класичного коефіцієнта детермінації R^2 як частки поясненої дисперсії.

© Publisher Bukovinian Math. Journal, 2026.
This is an Open Access article under the CC BY 4.0 license.
https://creativecommons.org/licenses/by/4.0/