How Good Is Your Synthetic Data?

Chattopadhyay $et.al$

***Abstract*—We propose a principled, model-agnostic method for evaluating the fidelity of synthetic tabular data based on whether conditional relationships among variables in the real date are preserved. In particular, we ask: Given a synthetic sample, if we resample a coordinate from the real data's conditional while holding the rest of the synthetic record fixed, does the realized value occur with high probability? This leads to a bounded, interpretable MAP-alignment statistic that directly measures how well a dataset reproduces its own learned conditional structure. Our approach induces a true metric on the space of underlying generative processes, with explicit finite-sample uncertainty bounds, and a one-sided fidelity score for assessing synthetic data against a trusted real dataset, based purely on samples, without invoking prior assumptions. Unlike moment-matching diagnostics such as covariance matrices, our approach evaluates the full conditional structure of the data, capturing nonlinear and higher-order dependencies, allowing us to precisely quantify with confidence: "how good is your synthetic data?"**

I. Introduction and Motivation

Synthetic tabular data are increasingly used for benchmarking, method development, privacy-preserving analysis, and regulatory reporting. Yet there remains no unified, principled answer to the basic question: *How similar is a synthetic dataset to the real one?* Current practices fall broadly into:

- **Likelihood-based metrics** ($e.g.$ ELBO [1], perplexity [2]), requiring the generator to have a tractable joint density;
- **Task-based metrics [3]**, such as training a classifier on synthetic data and testing on real data, which depend on subjective downstream tasks and can obscure structural mismatches;
- **Embedding-based metrics** such as FID [4] or classifier-based two-sample tests [5], which rely on feature extractors unrelated to the data domain and limit interpretability.

Another common but fragile practice is to compare covariance matrices of the real and synthetic datasets, along with component-wise means [6], [7]. This only captures pairwise linear relationships and can easily be matched by synthetic data that nevertheless distorts nonlinear, conditional, or higher-order structure. Later in the paper (Section VII), we explore two concrete examples in $\mathbb{R}^3$. In both cases, the "real" and "synthetic" datasets match *exactly* in all first- and second-order moments—including nontrivial covariance, but have very different higher order dependencies. The examples serve as tangible evidence for the necessity of evaluating synthetic data by its conditional structure rather than by low-order statistics.

In this paper we develop a complementary approach based on *conditional structure inferred directly from data*. Given samples (real or synthetic) we estimate each coordinate's conditional distribution $\phi^i(\cdot \mid x^{-i})$, $e.g.$ using reported conditional learners such as conditional inference trees. Then, for any sample x and coordinate i, we define a *MAP-alignment score* (notation: denotes x^i denotes the i^{th} variable, and x^{-i} denotes $\{x^j\}, j \neq i$):

$$\upsilon(x,i) = \frac{\phi^i(x^i \mid x^{-i})}{\max_y \phi^i(y \mid x^{-i})},$$

which has a direct generative interpretation:

Fix a synthetic record and consider one coordinate at a time. Hold all other coordinates fixed to their values in the synthetic sample, and imagine drawing the remaining coordinate from the real *data's conditional distribution. It then follows that $\upsilon(x,i)$ measures the odds of this sampling experiment to produce the observed value x_i.*

Averaging $\upsilon(x,i)$ over coordinates and samples in a given dataset D, yields a bounded, interpretable statistic $\Upsilon(D)$.

Thus, our statistic measures how often, and by how much, the observed data agree with conditional maximum-a-posteriori (MAP) [8] predictions. This approach is entirely *post hoc* and *model-agnostic*, and does not require access to the synthetic generator's internals. Under classical Brook-Dobrushin positivity conditions the full conditional system uniquely determines the joint distribution; connecting MAP-alignment to sharp identifiability results [9], [10], [11], [12] in the limit. We also show that our MAP-alignment profiles induce a distance between datasets that converges to a true metric on the space of underlying generative processes, with explicit finite-sample confidence bounds.

Compared to related work on synthetic data utility spaning global and task-based measures [13], [6], [7], [14], [15], [16], [17], which typically evaluate aggregate traits or downstream performance rather than the *record-level conditional structure*, our goal here is to quantify with confidence how similar is your synthetic data without invoking unvetted prior assumptions.

II. Formal Setup for Conditional Analysis

In our analysis we assume access to a procedure capable of estimating full conditional distributions from data, without committing to any specific method. In the applications (Section VIII) we instantiate this using a particular conditional learner, but the development that follows requires only that such (possibly noisy) conditional estimates can be obtained.

Asssume we hve a set of observable variables:

$$X = (X^1, \dots, X^N)$$

which take values in a finite product space

$$\mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^N.$$

Let P be the true data-generating distribution. For each coordinate i, define the full conditional

$$P_i(x^i \mid x^{-i}) := P(X^i = x^i \mid X^{-i} = x^{-i}).$$

We represent model conditional kernels as $\phi^i(\cdot \mid x^{-i})$. We assume strict positivity in the sense that, on the effective support of interest,

$$\phi^i(x^i \mid x^{-i}) > 0 \quad \text{for all } x,$$

and in practice enforce this by adding a small ε-floor to each conditional and renormalizing. This is the standard positivity assumption underlying Brook–Dobrushin factorization.

III. MAP-Alignment Functional

For any model $\{\phi^i\}$ and sample $x \in \mathcal{X}$, define

$$\upsilon(x,i) := \frac{\phi^i(x^i \mid x^{-i})}{\max_{y \in \mathcal{X}^i} \phi^i(y \mid x^{-i})},$$

which equals 1 exactly when x_i is a maximizer of the model conditional.

Given a dataset $D = \{x_k\}_{k=1}^M$, define

$$\Upsilon(D) := \frac{1}{MN} \sum_{k=1}^{M} \sum_{i=1}^{N} \upsilon(x_k, i),$$

an empirical estimate of

$$\Upsilon_\phi(P) = \mathbb{E}_{X \sim P}\left[\frac{1}{N}\sum_{i=1}^{N} \upsilon(X,i)\right].$$

A. Behavior Under Exact Conditionals

Lemma 1 (MAP-Alignment Under Exact Conditionals)**.** *Assume $\phi^i = P_i$ for all i. Fix i and x^{-i}, and let $p_j := P_i(j \mid x^{-i})$, $p_{\max} := \max_j p_j$. If $X^i \sim P_i(\cdot \mid x^{-i})$, then*

$$\upsilon(X,i) = \frac{P_i(X^i \mid x^{-i})}{p_{\max}}, \qquad \mathbb{E}[\upsilon(X,i) \mid x^{-i}] = \frac{1}{p_{\max}} \sum_j p_j^2.$$

Moreover:

- *$\upsilon(X,i) = 1$ iff $X_i \in \arg\max_j p_j$.*
- *If $P_i(\cdot \mid x^{-i})$ is uniform on its support, then $\mathbb{E}[\upsilon(X,i) \mid x^{-i}] = 1$.*
- *If some $p_j < p_{\max}$, then $\mathbb{E}[\upsilon(X,i) \mid x^{-i}] < 1$.*

Proof. Immediate from the definition and the fact that $\sum_j p_j^2 \le p_{\max} \sum_j p_j = p_{\max}$. □

Thus, under exact conditionals, $\upsilon(X,i)$ detects whether coordinate i is conditionally maximal. Define the oracle MAP-alignment:

$$\Upsilon_{\text{oracle}}(P) := \mathbb{E}_{X \sim P}\left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{X_i \in \arg\max_y P_i(y \mid X^{-i})\}\right].$$

IV. Brook–Dobrushin Factorization and Identifiability

Brook's lemma [9] and Dobrushin's consistency theorem [10] provide a complete characterization of joint distributions in terms of their full conditionals when strict positivity holds.

Assume

$$\phi^i(x^i \mid x^{-i}) = P_i(x^i \mid x^{-i}) \quad \text{whenever } P(x) > 0.$$

Choose a reference configuration x° with $P(x^\circ) > 0$. Define the interpolating sequence

$$x^{(0)} = x^\circ, \qquad x^{(i)} = (x^1, \dots, x^i, x^{i+1,\circ}, \dots, x^{N,\circ}),$$

so $x^{(N)} = x$. By repeated conditioning,

$$\frac{P(x^{(i)})}{P(x^{(i-1)})} = \frac{P_i(x^i \mid x^{<i}, x^{>i,\circ})}{P_i(x^{i,\circ} \mid x^{<i}, x^{>i,\circ})}.$$

Multiplying yields the Brook factorization [9]:

$$\frac{P(x)}{P(x^\circ)} = \prod_{i=1}^{N} \frac{P_i(x^i \mid x^{<i}, x^{>i,\circ})}{P_i(x^{i,\circ} \mid x^{<i}, x^{>i,\circ})}.$$

Replacing P_i by ϕ^i on the support gives an explicit reconstruction of P from the conditionals.

Theorem 1 (Nonparametric Brook–Dobrushin Identifiability)**.** *Assume:*

- *Strict positivity: $\phi^i(x^i \mid x^{-i}) > 0$ for all i and all x with $P(x) > 0$.*
- *Conditional accuracy: $\phi^i(x^i \mid x^{-i}) = P_i(x^i \mid x^{-i})$ whenever $P(x) > 0$.*

Then:

- *There exists a unique joint distribution $\widetilde{P}$ whose full conditionals are $\{\phi^i\}$.*
- *$\widetilde{P} = P$ on $\operatorname{supp}(P)$.*

Proof. Strict positivity implies uniqueness of a joint distribution compatible with the full conditionals [9], [10]. Since $\phi^i = P_i$ on the support, the reconstructed joint equals P there, and uniqueness forces equality. □

V. Evaluating Generators via Υ

For any model $\{\phi^i\}$,

$$\Upsilon_\phi(P) = \mathbb{E}_{X \sim P}\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\phi^i(X_i \mid X^{-i})}{\max_y \phi^i(y \mid X^{-i})}\right].$$

Let D_{test} be i.i.d. data from P. If $\phi_n^i \to P_i$ pointwise on the support and satisfy strict positivity, then dominated convergence yields

$$\Upsilon_{\phi_n}(P) \to \Upsilon_{\text{oracle}}(P).$$

Corollary 1 (Generator Convergence via MAP-Alignment)**.** *Let ϕ_n^i be strictly positive kernels converging to P_i on* $\operatorname{supp}(P)$. *Let $\widetilde{P}_n$ denote the joint compatible with $\{\phi_n^i\}$. Then:*

- *$\widetilde{P}_n \to P$ on $\operatorname{supp}(P)$.*
- *For any sequence of test sets D_{test} with $|D_{\text{test}}| \to \infty$, the empirical scores $\Upsilon(D_{\text{test}}; \phi_n)$ converge in probability to $\Upsilon_{\text{oracle}}(P)$.*

In particular, in regimes where the learned conditionals are known to converge to the truth, observing $\Upsilon(D_{\text{test}}; \phi_n)$ close to $\Upsilon_{\text{oracle}}(P)$ on sufficiently rich test data is strong evidence of accurate conditionals and, by Theorem 1, recovery of the joint law on the support. Conversely, systematically low MAP-alignment indicates mis-specified conditionals even when marginal summaries look satisfactory.

VI. Conditional Inference, MAP-Alignment Uncertainty, and a Metric on Underlying Processes

A. Inference of Conditionals Using Conditional Inference Trees

Let $D = \{x^k\}_{k=1}^n \subset \mathcal{X}^1 \times \cdots \times \mathcal{X}^N$ be a dataset in a finite product space. We construct a family of full conditional models $\Phi^{(n)} = \{\varphi^{i,n}(\cdot \mid x^{-i})\}_{i=1}^N$ by solving N supervised problems, one for each coordinate. For each i, a conditional inference tree is trained to predict X_i from X_{-i}, using an unbiased, permutation-based split selection and honest recursive partitioning as in [18], [19], [20]. Under standard regularity conditions (minimum node size $m_n \to \infty$, bounded depth growth, and intrinsic predictor dimension d_i), the conditional estimators satisfy the nonparametric uniform rate

$$\sup_{x_{-i}} \left|\varphi^{i,n}(x^i \mid x^{-i}) - P_i(x^i \mid x^{-i})\right| = O_p\left(n^{-1/(d_i+2)}\right), \quad (1)$$

where P_i denotes the true full conditional of the underlying generative process.

Strict positivity is imposed on the learned conditionals either intrinsically or via a small ε-floor. Under these assumptions, the Brook factorization and Dobrushin's uniqueness criterion [9], [10] guarantee that the full conditional family uniquely determines the compatible joint distribution, in the spirit of Hammersley–Clifford and Besag's analysis of Gibbs fields [11], [12]. Samples can then be generated by iteratively resampling missing coordinates from $\varphi_i^{(n)}(\cdot \mid x_{-i})$, while clamping observed coordinates when performing conditional inference.

B. MAP-Alignment and its Finite-Sample Uncertainty

Given a trained conditional family $\Phi^{(n)}$, define the MAP-alignment score of a datapoint x at coordinate i as

$$\upsilon_{\Phi^{(n)}}(x,i) = \frac{\varphi^{i,n}(x^i \mid x^{-i})}{\max_{y \in \mathcal{X}^i} \varphi^{i,n}(y \mid x_{-i})}, \quad (2)$$

which lies in $[0,1]$. For a test dataset $D_{\text{test}} = \{x_k\}_{k=1}^M$, define the empirical MAP-alignment functional

$$\hat{\Upsilon}_{\Phi^{(n)}}(D_{\text{test}}) = \frac{1}{MN} \sum_{k=1}^M \sum_{i=1}^N \upsilon_{\Phi^{(n)}}(x_k, i). \quad (3)$$

Let $Z = \upsilon_{\Phi^{(n)}}(X,i)$ where (X,i) is drawn uniformly from the test set and the index set. Since Z is bounded in $[0,1]$, Hoeffding's inequality [21] gives, for any $\varepsilon > 0$,

$$\Pr\left(\left|\hat{\Upsilon}_{\Phi^{(n)}}(D_{\text{test}}) - \mathbb{E}[Z]\right| > \varepsilon\right) \le 2\exp(-2MN\varepsilon^2). \quad (4)$$

Thus

$$\hat{\Upsilon}_{\Phi^{(n)}}(D_{\text{test}}) = \mathbb{E}[Z] + O_p\left((MN)^{-1/2}\right). \quad (5)$$

Combining this with the conditional estimation error in (1) yields

$$\hat{\Upsilon}_{\Phi^{(n)}}(D_{\text{test}}) = \Upsilon_P(D_{\text{test}}) + O_p\left((MN)^{-1/2} + n^{-1/(d_i+2)}\right), \quad (6)$$

providing full control of uncertainty from both the test sample and imperfect conditional learning.

C. A Metric on Underlying Generative Processes

For a strictly positive process P with full conditionals P_i, define its population υ-profile:

$$u_P(x,i) = \frac{P_i(x^i \mid x^{-i})}{\max_{y \in \mathcal{X}^i} P_i(y \mid x^{-i})}. \quad (7)$$

Fix a reference probability measure μ on $\mathcal{X} \times \{1,\dots,N\}$ whose support dominates all processes considered. Define the distance

$$d(P_1,P_2) = \mathbb{E}_{(x,i)\sim\mu}\left[|u_{P_1}(x,i) - u_{P_2}(x,i)|\right]. \quad (8)$$

Theorem 2. *Assume strict positivity and the Brook–Dobrushin uniqueness conditions [9], [10]. Then d is a metric on the space of generative processes compatible with these assumptions.*

Proof: Nonnegativity and symmetry are immediate. The triangle inequality follows from the scalar inequality $|a-c| \le |a-b| + |b-c|$ applied pointwise and integrated with respect to μ. For identity of indiscernibles, if $d(P_1,P_2) = 0$, then $u_{P_1}(x,i) = u_{P_2}(x,i)$ almost everywhere, implying equality of $P_{1,i}(\cdot \mid x_{-i})$ and $P_{2,i}(\cdot \mid x_{-i})$. Under strict positivity, the full conditionals uniquely determine the joint distribution by Brook–Dobrushin and Hammersley–Clifford [11], and therefore $P_1 = P_2$. ■

Given empirical datasets D_1, D_2, train conditional generator families G_1, G_2 yielding conditional families $\Phi^{(1)}$ and $\Phi^{(2)}$. Let $\hat{\mu}$ be the empirical distribution over $S = D_1 \cup D_2$. Define the empirical distance

$$\hat{d}(D_1,D_2) = \frac{1}{|S|N} \sum_{x \in S} \sum_{i=1}^N |u_{G_1}(x,i) - u_{G_2}(x,i)|. \quad (9)$$

Theorem 3. *Assume the conditional estimators satisfy* (1) *with errors ϵ_1, ϵ_2, and that the learned full conditionals are strictly positive on the effective support. Then*

$$\left|\hat{d}(D_1,D_2) - d(P_1,P_2)\right| = O_p\left((MN)^{-1/2} + \epsilon_1 + \epsilon_2\right).$$

Proof: The quantity $\hat{d} - d(P_1,P_2)$ admits the decomposition

$$|\hat{d} - d(G_1,G_2)| + |d(G_1,G_2) - d(P_1,P_2)|.$$

The first term is controlled by Hoeffding's inequality applied to the bounded random variables $|u_{G_1}(X,i) - u_{G_2}(X,i)|$, and the second by the Lipschitz continuity of $u(\varphi) = \varphi/\max_y \varphi(y)$ on strictly positive simplices, together with (1). This yields the stated rate. ■

D. One-Sided Fidelity: Evaluating Synthetic Data Against a Trusted Real Dataset

In many practical settings, the goal is not to compare two unknown processes symmetrically, but rather to assess how well a synthetic dataset D_{syn} approximates a trusted real dataset D_{real}. In such cases it is natural to evaluate D_{syn} using a *single* conditional system learned from D_{real}, rather than constructing separate conditional learners for both datasets.

Let $\Phi^{(\text{real})}$ be the conditional family fitted on D_{real} alone using any consistent conditional estimator (e.g., conditional inference trees). Define

$$\Upsilon_{\text{real}} = \hat{\Upsilon}_{\Phi^{(\text{real})}}(D_{\text{real}}), \qquad \Upsilon_{\text{syn}} = \hat{\Upsilon}_{\Phi^{(\text{real})}}(D_{\text{syn}}).$$

The difference

$$\Delta\Upsilon = \Upsilon_{\text{real}} - \Upsilon_{\text{syn}}$$

provides a natural *fidelity-to-real* score: it quantifies how well the synthetic records align with the conditional structure extracted from the real data. When D_{syn} is drawn from a process close to the real data-generating law P_{real}, we expect Υ_{syn} to be close to Υ_{real}. Conversely, structural distortions in the synthetic distribution manifest as systematic decreases in Υ_{syn}.

This one-sided measure is computationally simpler than the symmetric distance $d(P_1, P_2)$ and often of primary interest in applications where D_{real} is the reference dataset. The two-sided metric becomes essential when comparing two arbitrary datasets on equal footing, or when estimating the true distance between underlying generative processes.

E. Consistency of the Dataset Distance as a Metric on Processes

Theorem 4 (Convergence to the Population Metric)**.** *Let D_1 and D_2 be drawn i.i.d. from strictly positive processes P_1 and P_2. Suppose the conditional generators satisfy* $\sup_{x^{-i}} |\varphi^{i,k} - P_{k,i}| \le \epsilon_k$ *with $\epsilon_k \to 0$ in probability as $|D_k| \to \infty$, and suppose the test-set size $M \to \infty$. Then*

$$\hat{d}(D_1, D_2) \xrightarrow{p} d(P_1, P_2). \tag{10}$$

Proof: From the previous theorem we have

$$\left|\hat{d}(D_1, D_2) - d(P_1, P_2)\right| = O_p\Big((MN)^{-1/2} + \epsilon_1 + \epsilon_2\Big).$$

As $M \to \infty$ and $|D_k| \to \infty$, both $(MN)^{-1/2}$ and ϵ_k converge to zero in probability, which implies (10). ■

Corollary 2. *If $P_1 = P_2$, then $\hat{d}(D_1, D_2) \xrightarrow{p} 0$. If $P_1 \neq P_2$ differ in their full conditional structure on any set of positive μ-measure, then $\hat{d}(D_1, D_2) \xrightarrow{p} d(P_1, P_2) > 0$.*

F. Pseudocode for Estimating the MAP-Alignment Distance

To make the computation explicit, Algorithm 3 summarizes the empirical distance estimation between two datasets D_1 and D_2 via their trained conditional generator families G_1 and G_2.

G. Finite-Sample Uncertainty and Numerical Examples

For confidence level $1-\delta$, Hoeffding's inequality [21] yields

$$|\hat{d}(D_1, D_2) - d(G_1, G_2)| \le \sqrt{\frac{1}{2MN} \log\frac{2}{\delta}}$$

with probability at least $1 - \delta$. Adding conditional estimation error gives

$$|\hat{d}(D_1, D_2) - d(P_1, P_2)| \le \sqrt{\frac{1}{2MN} \log\frac{2}{\delta}} + (\epsilon_1 + \epsilon_2).$$

Algorithm 1: Computation of the MAP-alignment functional υ and aggregate score $\Upsilon(D_{\text{test}})$

```
Input  :  Test dataset $D_{\text{test}} = \{x^{(k)}\}_{k=1}^{M}$,
          Conditional kernels $\{\varphi_i(\cdot \mid x^{-i})\}_{i=1}^{N}$,
          Finite state spaces $\{\mathcal{X}^i\}_{i=1}^{N}$ for each
coordinate.
Output:  Per-sample MAP-alignment matrix $\upsilon_i^{(k)}$,
          Aggregate MAP-alignment score $\Upsilon(D_{\text{test}})$.
Initialize $\Upsilon \leftarrow 0$
for $k \leftarrow 1$ to $M$ do
    for $i \leftarrow 1$ to $N$ do
        Let $x_i^{(k)}$ be the $i$-th coordinate of sample $x^{(k)}$
        Let $x^{(k),-i}$ be all coordinates of $x^{(k)}$ except $i$
        // Compute model conditional for
           the observed value
        $p_{\text{obs}} \leftarrow \varphi_i\big(x_i^{(k)} \mid x^{(k),-i}\big)$
        // Compute maximum conditional
           probability over the state
           space
        $p_{\max} \leftarrow 0$
        foreach $y \in \mathcal{X}^i$ do
            $p_y \leftarrow \varphi_i\big(y \mid x^{(k),-i}\big)$
            if $p_y > p_{\max}$ then
                $p_{\max} \leftarrow p_y$
        // MAP-alignment for sample $k$,
           coordinate $i$
        $\upsilon_i^{(k)} \leftarrow \dfrac{p_{\text{obs}}}{p_{\max}}$
        // Accumulate for the aggregate
           score
        $\Upsilon \leftarrow \Upsilon + \upsilon_i^{(k)}$
// Normalize by number of samples and
   coordinates
$\Upsilon(D_{\text{test}}) \leftarrow \dfrac{\Upsilon}{M \cdot N}$
return $\{\upsilon_i^{(k)}\}_{k=1,\ldots,M}^{i=1,\ldots,N}$, $\Upsilon(D_{\text{test}})$
```

Example 1 (Moderate dataset). Let $N = 100$, $M = 100$, $\epsilon_1 = \epsilon_2 = 0.05$, and $\delta = 0.05$. Then

$$\sqrt{\frac{1}{2MN}\log(40)} \approx 0.0136, \qquad \epsilon_1 + \epsilon_2 = 0.10,$$

so the total uncertainty radius is approximately 0.1136. An observed value $\hat{d}(D_1, D_2) = 0.40$ is therefore consistent with a true distance $d(P_1, P_2)$ lying roughly in $[0.29, 0.51]$.

Example 2 (Large test set, improved conditionals). Let $N = 100$, $M = 1000$, $\epsilon_1 = \epsilon_2 = 0.02$, and $\delta = 0.05$. Then

$$\sqrt{\frac{1}{2MN}\log(40)} \approx 0.0043, \qquad \epsilon_1 + \epsilon_2 = 0.04,$$

giving a total bound near 0.0443. An empirical distance $\hat{d}(D_1, D_2) = 0.20$ then implies $d(P_1, P_2)$ is concentrated in the interval $[0.156, 0.244]$.

These values show that the MAP-alignment geometry yields a statistically well-controlled and data-efficient method for

Algorithm 2: One-Sided MAP-Alignment Fidelity to a Trusted Real Dataset

```
Input  :  Real dataset $D_{\text{real}}$;
          Synthetic dataset $D_{\text{syn}}$;
          Conditional learner $\mathsf{LearnConditionals}$;
          Significance level $\alpha \in (0,1)$.
Output:  Estimated real MAP-alignment $\hat{\Upsilon}_{\text{real}}$;
          Estimated synthetic MAP-alignment $\hat{\Upsilon}_{\text{syn}}$;
          Fidelity score $\Delta\hat{\Upsilon}$ and $(1-\alpha)$ confidence interval.
Fit conditional family on real data
  $\Phi^{(\text{real})} \leftarrow \mathsf{LearnConditionals}(D_{\text{real}})$
Compute MAP-alignment on real data
  $M_{\text{real}} \leftarrow |D_{\text{real}}|$
  $A_{\text{real}} \leftarrow 0$
  foreach $x \in D_{\text{real}}$ do
      for $i \leftarrow 1$ to $N$ do
          Compute $\varphi_i(x_i \mid x_{-i})$ from $\Phi^{(\text{real})}$
          Compute $m_i(x) \leftarrow \max_{y\in\mathcal{X}^i} \varphi_i(y \mid x_{-i})$
          Compute $\upsilon(x,i) \leftarrow \varphi_i(x_i \mid x_{-i})/m_i(x)$
          $A_{\text{real}} \leftarrow A_{\text{real}} + \upsilon(x,i)$
  $\hat{\Upsilon}_{\text{real}} \leftarrow A_{\text{real}}/(M_{\text{real}}N)$
Compute MAP-alignment on synthetic data using the same conditionals
  $M_{\text{syn}} \leftarrow |D_{\text{syn}}|$
  $A_{\text{syn}} \leftarrow 0$
  foreach $x \in D_{\text{syn}}$ do
      for $i \leftarrow 1$ to $N$ do
          Compute $\varphi_i(x_i \mid x_{-i})$ from $\Phi^{(\text{real})}$
          Compute $m_i(x) \leftarrow \max_{y\in\mathcal{X}^i} \varphi_i(y \mid x_{-i})$
          Compute $\upsilon(x,i) \leftarrow \varphi_i(x_i \mid x_{-i})/m_i(x)$
          $A_{\text{syn}} \leftarrow A_{\text{syn}} + \upsilon(x,i)$
  $\hat{\Upsilon}_{\text{syn}} \leftarrow A_{\text{syn}}/(M_{\text{syn}}N)$
Compute fidelity score
$\Delta\hat{\Upsilon} \leftarrow \hat{\Upsilon}_{\text{real}} - \hat{\Upsilon}_{\text{syn}}$
Compute Hoeffding-based confidence radius for $\Delta\hat{\Upsilon}$
$r_{\text{real}} \leftarrow \sqrt{\frac{1}{2M_{\text{real}}N}\log\left(\frac{4}{\alpha}\right)}$
$r_{\text{syn}} \leftarrow \sqrt{\frac{1}{2M_{\text{syn}}N}\log\left(\frac{4}{\alpha}\right)}$
$r_\Delta \leftarrow r_{\text{real}} + r_{\text{syn}}$
Construct $(1-\alpha)$ confidence interval (clipped to $[-1,1]$)
$L \leftarrow \max\{-1, \Delta\hat{\Upsilon} - r_\Delta\}$
$U \leftarrow \min\{1, \Delta\hat{\Upsilon} + r_\Delta\}$
return $\hat{\Upsilon}_{\text{real}}, \hat{\Upsilon}_{\text{syn}}, \Delta\hat{\Upsilon}, [L,U]$
```

Algorithm 3: Estimation of MAP-alignment distance and confidence interval

```
Input  :  Datasets $D_1$ and $D_2$;
          Trained conditional generator families $G_1$ and $G_2$ with conditional families $\Phi^{(1)}$ and $\Phi^{(2)}$;
          Number of variables $N$;
          Significance level $\alpha \in (0,1)$ (for a $(1-\alpha)$ confidence interval);
          Bounds on conditional estimation error $\epsilon_1, \epsilon_2$ for $G_1, G_2$.
Output:  Empirical distance $\hat{d}(D_1,D_2)$ and $(1-\alpha)$
          confidence interval $[L,U]$ for $d(P_1,P_2)$.
Construct the pooled set $S \leftarrow D_1 \cup D_2$
Initialize accumulator $A \leftarrow 0$
Let $M \leftarrow |S|$
foreach $x \in S$ do
    for $i \leftarrow 1$ to $N$ do
        Compute $\upsilon_1 \leftarrow \upsilon_{\Phi^{(1)}}(x,i)$ using (2)
        Compute $\upsilon_2 \leftarrow \upsilon_{\Phi^{(2)}}(x,i)$ using (2)
        Set $A \leftarrow A + |\upsilon_1 - \upsilon_2|$
Set $\hat{d}(D_1,D_2) \leftarrow A/(MN)$
Compute the Hoeffding test-noise radius
        $r_{\text{test}} \leftarrow \sqrt{\frac{1}{2MN}\log\left(\frac{2}{\alpha}\right)}$
Set the total radius
        $r_{\text{total}} \leftarrow r_{\text{test}} + (\epsilon_1 + \epsilon_2)$
Compute confidence interval bounds
        $L \leftarrow \max\{0,\ \hat{d}(D_1,D_2) - r_{\text{total}}\}$
        $U \leftarrow \min\{1,\ \hat{d}(D_1,D_2) + r_{\text{total}}\}$
return $\hat{d}(D_1,D_2)$, $L$, $U$
```

distinguishing whether two datasets arise from the same underlying conditional generators.

VII. Low-Order Moment Matching is Insufficient: Two Illustrative Examples

Comparing columnwise means, variances, or covariance matrices is a common practice for evaluating synthetic tabular data. However, agreement in these low-order summaries does *not* imply that two datasets share similar joint or conditional structure. We present two examples in $\mathbb{R}^3$ that highlight this gap: in both cases, the synthetic and real datasets match exactly in all first- and second-order moments, yet differ substantially in their conditional behavior. In each case, the MAP-alignment statistic υ reveals discrepancies that covariance structures alone cannot detect.

A. Example 1: Uniform vs. Gaussian with Identical Moments

Consider the following two distributions on $\mathbb{R}^3$:

$$X^{(U)} \sim \text{Uniform}([-\sqrt{3},\sqrt{3}]^3), \qquad X^{(G)} \sim \mathcal{N}_3(0, I_3).$$

Both satisfy

$$\mathbb{E}[X^{(U)}] = \mathbb{E}[X^{(G)}] = 0, \qquad \text{Cov}(X^{(U)}) = \text{Cov}(X^{(G)}) = I_3.$$

Hence all marginal means, variances, and the full covariance matrix coincide.

Yet their conditional structures differ sharply. For $X^{(U)}$, each coordinate has a flat conditional density on $[-\sqrt{3},\sqrt{3}]$, yielding

$$\upsilon(X^{(U)}, i) = 1 \quad \text{for almost every sample.}$$

For $X^{(G)}$, the ith coordinate conditional is $X_i^{(G)} \mid X_{-i}^{(G)} \sim \mathcal{N}(0,1)$, giving

$$\upsilon(X^{(G)}, i) = \exp\left(-\tfrac{1}{2}X_i^{(G)2}\right),$$

with mean approximately 0.71. Thus, although the datasets match in mean and covariance, the MAP-alignment values differ substantially:

$$\Upsilon(X^{(U)}) \approx 1, \qquad \Upsilon(X^{(G)}) \approx 0.7.$$

Low-order moments fail to detect this discrepancy.

B. Example 2: Matching Non-Identity Covariances with Distinct Conditionally

To demonstrate that the limitation persists even when the covariance matrix is nontrivial, apply the same invertible linear map

$$L = \begin{pmatrix} 1 & \rho & 0 \\ 0 & 1 & \rho \\ 0 & 0 & 1 \end{pmatrix}, \qquad 0 < \rho < 1,$$

to both datasets and define

$$Y^{(U)} = LX^{(U)}, \qquad Y^{(G)} = LX^{(G)}.$$

Both transformed datasets have the *same* non-identity covariance matrix

$$\mathrm{Cov}(Y^{(U)}) = \mathrm{Cov}(Y^{(G)}) = \Sigma = LL^\top,$$

and share identical columnwise means and variances.

However, their conditional distributions remain fundamentally different. Since $Y^{(U)}$ is the image of a uniform cube under a linear shear, each conditional $Y^{i,(U)} \mid Y^{-i,(U)}$ is uniform on a finite interval (given by the intersection of a line with a parallelepiped), implying

$$\upsilon(Y^{(U)}, i) \approx 1.$$

Conversely, $Y^{(G)} \sim \mathcal{N}_3(0, \Sigma)$ has linear–Gaussian conditionals. If $Y^{i,(G)} \mid Y^{-i,(G)} \sim \mathcal{N}(m_i(y^{-i}), \sigma_i^2)$, then

$$\upsilon(Y^{(G)}, i) = \exp\left(-\tfrac{1}{2}(y^i - m_i(y^{-i}))^2/\sigma_i^2\right),$$

with empirical means typically in the range 0.7–0.8.

Thus, even though $Y^{(U)}$ and $Y^{(G)}$ agree in all first- and second-order statistics, their conditional behavior differs markedly, and the MAP-alignment statistic again reveals the discrepancy:

$$\Upsilon(Y^{(U)}) \approx 1, \qquad \Upsilon(Y^{(G)}) \approx 0.7\text{–}0.8.$$

C. Implications

These examples demonstrate that matching means, variances, or full covariance matrices (even with non-identity structure) is insufficient to conclude that two datasets arise from similar generative processes. Higher-order structure, multimodality, support geometry, and conditional relationships can remain entirely undetected by low-order moments. MAP-alignment, by directly assessing conditional fidelity, exposes such differences immediately.

VIII. Applications: Comparing Synthesizers

References

[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. International Conference on Learning Representations (ICLR)*, 2014, arXiv:1312.6114.

[2] F. Jelinek, R. L. Mercer, L. R. Bahl, and J. K. Baker, "Perplexity—a measure of the difficulty of speech recognition tasks," *Journal of the Acoustical Society of America*, 1977, short paper introducing perplexity for language models.

[3] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs," *arXiv preprint arXiv:1706.02633*, 2017.

[4] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, introduces the Fréchet Inception Distance (FID).

[5] D. Lopez-Paz and M. Oquab, "Revisiting classifier two-sample tests," in *Proc. International Conference on Learning Representations (ICLR)*, 2017.

[6] J. Snoke, G. M. Raab, B. Nowok, C. Dibben, and A. Slavkovic, "General and specific utility measures for synthetic data," *Journal of Privacy and Confidentiality*, vol. 8, no. 1, 2018.

[7] B. Nowok, G. M. Raab, and C. Dibben, "synthpop: Bespoke creation of synthetic data in R," *Journal of Statistical Software*, vol. 74, no. 11, 2016.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[9] D. Brook, "On the distinction between conditional and unconditional distributions," *Biometrika*, vol. 51, no. 3–4, pp. 481–483, 1964.

[10] R. L. Dobrushin, "The description of a random field by means of conditional probabilities and conditions of its regularity," *Theory of Probability and its Applications*, vol. 13, no. 2, pp. 197–224, 1968.

[11] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," in *Markov Random Fields*, P. Grimmett, Ed. Springer, 2017, originally written in 1971 as an unpublished manuscript.

[12] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, vol. 36, no. 2, pp. 192–236, 1974.

[13] M.-J. Woo, J. P. Reiter, A. Oganian, and A. F. Karr, "Global measures of data utility for microdata masked for disclosure limitation," *Journal of Privacy and Confidentiality*, vol. 1, no. 1, pp. 111–124, 2009.

[14] K. El Emam, "Seven ways to evaluate the utility of synthetic data," *IEEE Security & Privacy*, vol. 18, no. 4, pp. 56–59, 2020.

[15] K. El Emam, L. Mosquera, X. Fang, and A. El-Hussuna, "Utility metrics for evaluating synthetic health data generation methods: Validation study," *JMIR Medical Informatics*, vol. 10, no. 4, p. e35734, 2022.

[16] F. K. Dankar, M. K. Ibrahim, and L. Ismail, "A multi-dimensional evaluation of synthetic data generators," *IEEE Access*, vol. 10, pp. 11 147–11 158, 2022.

[17] B. Kaabachi, J. Despraz, T. Meurers, K. Otte, M. Halilovic, B. Kulynych, F. Prasser, and J. L. Raisaro, "A scoping review of privacy and utility metrics in medical synthetic data," *npj Digital Medicine*, vol. 8, no. 1, p. 60, 2025.

[18] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 651–674, 2006.

[19] ——, "party: A laboratory for recursive partytioning," *R News*, vol. 6, no. 2, pp. 17–23, 2006.

[20] C. Strobl, A. Boulesteix, T. Hothorn, and A. Zeileis, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformatics*, vol. 8, no. 25, pp. 1–21, 2007.

[21] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.