Equivalence of several curves assessing the similarity between probability distributions

Loic Simon  
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC  
loic.simon@ensicaen.fr

Julien Rabin  
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC

Ryan Webster  
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC

Abstract

The recent advent of powerful generative models has triggered the renewed development of quantitative measures to assess the proximity of two probability distributions. As the scalar Frechet inception distance remains popular, several methods have explored computing entire curves, which reveal the trade-off between the fidelity and variability of the first distribution with respect to the second one. Several of such variants have been proposed independently and while intuitively similar, their relationship has not yet been made explicit. In an effort to make the emerging picture of generative evaluation more clear, we propose a unification of four curves known respectively as: the precision-recall (PR) curve, the Lorenz curve, the receiver operating characteristic (ROC) curve and a special case of Rényi divergence frontiers.

1 Introduction

Generative models, particularly Generative Adversarial Networks (GANs), are already in an uncanny valley of realism. Indeed, quality of generation has improved significantly to the point where for certain datasets, human observers have difficulty discerning real and fake [1][2]. As these networks see real application, evaluation of generative networks has become essential and remains challenging. For instance, the privacy and security of generative models has become paramount, including the largest ever kaggle competition to address deep fakes, or several new works addressing how generative models leak training data [3][4]. Even properly determining sample quality remains challenging [5]. The Frechet Inception Distance (FID) [6] was shown to correlate decently with human evaluation and remains the most popular evaluation metric, but as a scalar metric is limited when assessing model failure [7]. A variety of other approaches attempt to give an empirical estimation of sample quality, for instance in [8], the original GAN training divergence was re-used for evaluation. [7] proposed computing an entire precision-recall curve for the generated distribution. Unlike the scalar FID, this curve distinguishes what we’ll refer to as the fidelity and variability of the model [9]. Fidelity evaluates whether the generated distribution produces data that are faithful to the original distribution whereas variability reflects the fact that it covers the entire distribution with the correct importance. For instance, a generator of facial images with poor diversity may only generate one gender, whereas a generator with poor fidelity contains generation artefacts.
Of course, estimating the similarity of two distributions is not a new problem. Many other probability
distribution metrics can be used such as divergences \[10\] or integral probability metrics \[11, 12\] to
name a few. In this work, we’ll take special focus on the recent work of \[7\] computing a precision-
recall type curve between two distributions. Several works explicitly build upon their definition and
propose some extensions. For instance, \[13\] generalizes the PR-curve to arbitrary probability (while
the work of \[7\] was restricted to a discrete settings). More practical works aim at improving empirical
evaluation of fidelity and variability \[14, 9\]. \[15\] proposed the (Renyi) divergence frontiers; this
alternate curve coincides with the original precision and recall curve (PRC) for discrete distributions
when the Renyi exponent is infinite.

Independently, a handful of alternative curves were defined to compare two distributions. For instance,
ROC curves were proposed by \[16, 17\] and the Lorenz curves by \[18, 19\]. Despite their purportedly
disparate definitions, these alternate notions are tightly linked with the PR curves. A diagram of their
interconnections is displayed in Fig 1 which constitute the subject of this work. Our contribution
is the theoretical unification between the involved curves. We first consider the link found by \[15\]
between PR curves and divergence frontiers for infinite Renyi exponent, hence extending it from
discrete distributions to general ones. Second, we show that ROC and Lorenz curves are essentially
the same thing. Last, we show that Lorenz curves and PR curves are related through convex duality.
A few simple toy examples are considered in Appendix B to expose the different properties of Lorenz
and precision-recall curves as well as to highlight their dual nature.

2 Notions from standard measure theory

We start these notes by recalling some standard notations, definitions, and results of measure theory.
For the remainder, \((\Omega, A)\) represents a common measurable space, and we will denote \(\mathcal{M}(\Omega)\) the
set of signed measures, \(\mathcal{M}^+(\Omega)\) the set of positive measures and \(\mathcal{M}_p(\Omega)\) the set of probability
distributions over that measurable space. The extended half real-line is denoted by \(\mathbb{R}^+ = \mathbb{R}^+ \cup \{\infty\}\).

**Definition 1.** Let \(\mu, \nu\) two signed measures. We denote by \(\text{supp}(\mu)\) the support of \(\mu\), \(|\mu|\) the total
variation measure of \(\mu\), \(\frac{d\mu}{d\nu}\) the Radon-Nikodym derivative of \(\mu\) w.r.t. \(\nu\) and \(\mu \wedge \nu = \min(\mu, \nu) := \frac{1}{2}(\mu + \nu - |\mu - \nu|)\) (a.k.a. the measure of largest common mass between \(\mu\) and \(\nu\) \[20\]). Besides, as
usual, \(\mu \ll \nu\) means that \(\mu\) is absolutely continuous w.r.t. \(\nu\).

Later on, we shall rely on the following general result.

**Proposition 1.** Let \(\mu \in \mathcal{M}_p(\Omega)\) and \(\nu \in \mathcal{M}_p(\Omega)\) two distributions.

\[\text{To avoid technical problems (e.g. to ensure the existence and uniqueness of Radon-Nikodym derivatives), all measures are supposed to be } \sigma\text{-finite.}\]
Another useful result of \cite{13} is summarized now:

**Theorem 2.** Let

\[ \frac{\mu(A)}{\nu(A)} = \text{ess sup}_\mu \frac{d\mu}{d\nu}; \]

1. If \( \mu \ll \nu \) then \( \sup_{A \in \mathcal{A}} \frac{\mu(A)}{\nu(A)} = \text{ess sup}_\mu \frac{d\mu}{d\nu}; \)

2. otherwise, \( \sup_{A \in \mathcal{A}} \frac{\mu(A)}{\nu(A)} = \text{ess sup}_\mu \frac{d\mu}{d\nu}/\frac{d\nu}{d\mu}. \)

**Proof.** For the sake of completeness, we provide a proof of this technical result in Appendix A. \( \square \)

### 3 Several curve alternatives to assess distributional closeness

In this section, we review several curves proposed in the literature to assess the similarity between two distributions \( P \) and \( Q \). We simply recap the principal definitions and useful results. Some notions are subject to minor adaptations in order to simplify the exposition of the links between the considered curves. Anytime such a revision is adopted, it shall be explicitly mentioned.

#### 3.1 Precision-recall curves

The precision-recall curves were first proposed by \cite{7} for discrete distributions and then extended to the general case by \cite{13}. We follow the definition of the latter up to a minor fix.\(^3\)

**Definition 2.** Let \( P, Q \) two distributions from \( \mathcal{M}_p(\Omega) \). We refer to the Precision-Recall set \( \text{PRD}(P, Q) \) as the set of Precision-Recall pairs \( (\alpha, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+ \) such that

\[ \exists \mu \in \text{AC}(P, Q), P \geq \beta \mu, Q \geq \alpha \mu, \]

where \( \text{AC}(P, Q) := \{ \mu \in \mathcal{M}_p(\Omega)/\mu \ll P \text{ and } \mu \ll Q \}. \)

The precision value \( \alpha \) is related to the proportion of the generated distribution \( Q \) that matches the true data \( P \), while conversely the recall value \( \beta \) is the amount of the distribution \( P \) that can be reconstructed from \( Q \). Because of the lack of natural order on \([0,1] \times [0,1]\), \cite{13} has proposed to focus on the Pareto front of \( \text{PRD}(P, Q) \) defined as follows:

**Definition 3.** The precision recall-curve \( \partial \text{PRD}(P, Q) \) is the set of \( (\alpha, \beta) \in \text{PRD}(P, Q) \) such that

\[ \forall (\alpha', \beta') \in \text{PRD}(P, Q), \alpha \geq \alpha' \text{ or } \beta \geq \beta'. \]

In fact, this frontier is a curve for which \cite{7} have exposed a parameterization, later generalized by \cite{13}. We recall their result now.

**Theorem 1.** Let \( P, Q \) two distributions from \( \mathcal{M}_p(\Omega) \) and \( (\alpha, \beta) \) non negative. Then, denoting\(^4\)

\[ \forall \lambda \in \mathbb{R}^+, \begin{cases} \alpha_\lambda := (\lambda P) \land Q \land (\Omega) \\ \beta_\lambda := (P \land Q) \land (\Omega) \end{cases} \]

1. \( (\alpha, \beta) \in \text{PRD}(P, Q) \) iff \( \alpha \leq \alpha_\lambda \text{ and } \beta \leq \beta_\lambda \) where \( \lambda := \frac{\alpha}{\beta} \in \mathbb{R}^+ \).

2. As a result, the PR curve can be parameterized as:

\[ \partial \text{PRD}(P, Q) = \{ (\alpha_\lambda, \beta_\lambda)/\lambda \in \mathbb{R}^+ \}. \]

Another useful result of \cite{13} is summarized now:

**Theorem 2.** Let \( P, Q \) two distributions from \( \mathcal{M}_p(\Omega) \). Then

\[ \forall \lambda \in \mathbb{R}^+, \begin{cases} \alpha_\lambda = \lambda \left(1 - P(A_{\lambda}^{Q/P})\right) + Q(A_{\lambda}^{Q/P}) \\ \beta_\lambda = 1 - P(A_{\lambda}^{Q/P}) + \frac{Q(A_{\lambda}^{Q/P})}{\lambda} \end{cases}, \]

where the likelihood ratio sets are defined as

\[ A_{\lambda}^{Q/P} := \{ \frac{dQ}{dP+Q} \leq \lambda \frac{dP}{dP+Q} \}. \]

\(^3\)There is an issue with their original definition where \((1,0)\) and \((0,1)\) are always in \( \text{PRD}(P, Q) \), while they should not when part of the mass of \( P \) is absent from \( Q \) and vice versa. Our fix consists in considering only the distributions \( \mu \) that are absolutely continuous w.r.t \( P \) and \( Q \).

\(^4\)As is conventionally surmised in measure theory \( 0 \times \infty = 0 \) so that \( \alpha_\infty = Q(\text{supp}(P)) \) and \( \beta_0 = P(\text{supp}(Q)) \).
3.2 Divergence frontiers

Divergence frontiers were proposed very recently by [15] as a generalization of precision-recall curves. Such a notion builds upon the Renyi divergence between two distributions.

**Definition 4.** Let \( \mu, \nu \in \mathcal{M}_\mu(\Omega) \) two distributions such that \( \mu \ll \nu \) and \( a \in \mathbb{R}^+ \setminus \{1\} \). The \( a \)-Renyi-divergence between \( \mu \) and \( \nu \) is defined as:

\[
D_a(\mu \parallel \nu) := \log \left( \left\| \frac{d\mu}{d\nu} \right\|_{a-1,d\mu} \right)
\]

(6)

where the invoked norm is defined as

\[
\| f \|_{a-1,d\mu} := \left( \int f^{a-1} d\mu \right)^\frac{1}{a-1}
\]

when \( a < +\infty \) and is the essential supremum norm for \( a = \infty \). Besides when \( a = 1 \), this definition is extended by continuity and leads to the KL-divergence.

We adapt the definition of divergence frontiers from [15].

**Definition 5 (Divergence frontiers).** Let \( P, Q \) two distributions and \( a \in \mathbb{R}^+ \). Then the exclusive realizable divergence region is defined as the set:

\[
\mathcal{R}_a^0(P, Q) := \{(D_a(\mu \parallel Q), D_a(\mu \parallel P)) / \mu \in \text{AC}(P, Q)\}
\]

(7)

And the exclusive divergence frontier is defined as the (weak) Pareto front of this region, that is the set \( \partial \mathcal{R}_a^0 \) of couples \( (\pi, \rho) \in \mathcal{R}_a^0 \) such that:

\[
\forall (\pi', \rho') \in \mathcal{R}_a^0, \pi \leq \pi' \text{ or } \rho \leq \rho'.
\]

(8)

In the event that \( \mathcal{R}_a^0(P, Q) = \emptyset \), the frontier is by convention restricted to the point \((+\infty, +\infty)\).

3.3 Lorenz and ROC curves

Lorenz curves were originally introduced by [21] to delineate income inequalities. In essence they highlight how much a single one dimensional distribution differs from a uniform distribution. This notion was then generalized to characterize the closeness of two arbitrary distributions by [18] [19].

**Definition 6.** Let \( P, Q \) two distributions from \( \mathcal{M}_\mu(\Omega) \). One defines the Lorenz diagram between \( P \) and \( Q \) as

\[
LD(P, Q) = \left\{ \left( \int f dP, \int f dQ \right) / 0 \leq f \leq 1 \right\},
\]

(9)

where the function \( f \) is required to be measurable.

Then the Lorenz curve between \( P \) and \( Q \) is defined as the lower envelop of the Lorenz diagram:

\[
F_{LD}^{P,Q}(t) := \inf_{0 \leq f \leq 1} \int f dQ.
\]

(10)

In absence of ambiguity on the involved distributions, we shall denote it simply \( F(t) \) rather than \( F_{LD}^{P,Q}(t) \). This curve is easily shown to be a monotonic and convex function.

**Remark 1.** If one considers in (9) only the range of indicator functions, then one recovers a subset of the Lorenz diagram, from which the Lorenz diagram can be extracted by merely taking the closed convex hull. This fact underpins the equivalence between Lorenz diagrams/curves and the seemingly different notions of Mode Collapse Region / ROC curves proposed by [16]. Indeed, they show in [17] Remark 6) that their Mode Collapse Region (MCR) can be obtained as the convex hull of the set of points \( (P(A), Q(A)) \) where \( A \) is any measurable set such that \( Q(A) \geq P(A) \). As such the MCR is the upper half of the Lorenz diagram when one cuts it along the main diagonal[4]. Then the authors proceed to define the ROC curve as the upper envelop of the MCR, which in turns is the symmetric transform of the lower envelop (i.e. the Lorenz curve) along the same diagonal. For the sake of time precedence, we shall thereafter focus solely on the Lorenz diagram and curve.

[4]In the original work, the distribution \( \mu \) may range on a restricted set of distributions such as an exponential family. Besides, to avoid situations where the divergence is ill-defined, we imposed that \( \mu \) is absolutely continuous w.r.t both \( P \) and \( Q \).

[5]That is to say, the line segment between \((0, 0)\) and \((1, 1)\).
4 Links between the alternative curves

4.1 Short preamble

Before going into further details about how the aforementioned curves relate to one another, let us provide a few general facts about how they differ. To make our discussion more concrete, we consider a simple illustrative case in Fig 2 where P and Q are two mixtures of Gaussian. One can first note that each curve is subject to specific “regularity” properties such as convexity and boundedness. For instance contrary to the Lorenz curve, the PR curve does not enjoy any convexity property. Similarly, both the PR curve and the Lorenz curve are bounded within the domain $[0, 1] \times [0, 1]$, while the divergence frontiers are not bounded in general.

But, let us set aside any such consideration and focus instead on the principal function of the three curves under consideration: namely how they characterize the similarity between P and Q. For a given curve alternative, one may consider two extreme configurations. One one hand, the perfect match between P and Q i.e. $P = Q$ is represented in dashed blue. On the other hand, the complete discord between P and Q, denoted by $P \perp Q$ corresponds to an empty overlap of their supports (or more formally to two mutually singular distributions) and is represented in green.

Then a particular instance of the considered curve will appear as an in-between case. The closer it stands to the blue spot (and therefore the farther from the green spot), the more similar P and Q. The interested reader can find a few concrete comparison examples in Appendix B.

4.2 Precision-recall vs divergence frontiers

In [15], it is shown that in the case of discrete measures, the notion of divergence frontier matches precision-recall curve in the limit case where the Renyi exponent $\alpha \to \infty$. We extend here this result to the case of general distributions.

Note that for the divergence frontier, $P \perp Q \implies \mathcal{R}^\alpha_{\text{div}}(P, Q) = \emptyset$ and by convention the divergence frontier is $(+\infty, +\infty)$ which is why the green location does not appear in the illustration.
Theorem 3. Let $P, Q$ two distributions. Then,
$$\partial PRD(P, Q) = \{(e^{-\pi}, e^{-\rho})/(\pi, \rho) \in \partial R^+_{\infty}(P, Q)\}$$

Proof. From the definition of the Renyi divergence it is clear that
$$D_{\infty}(\mu \parallel Q) = \log \left( \sup_{\lambda \in R} \frac{\mu(A)}{Q(A)} \right).$$
Similarly,
$$D_{\infty}(\mu \parallel P) = \log \left( \sup_{\lambda \in R} \frac{\mu(A)}{P(A)} \right).$$
Besides it is clear that the precision-recall curve is obtained as the Pareto-front of the set:
$$\left\{ \left( \inf_{A \in A, \mu(A) > 0} \frac{Q(A)}{\mu(A)}, \inf_{A \in A, \mu(A) > 0} \frac{P(A)}{\mu(A)} \right) / \mu \in AC(P, Q), \mu \in M_p(\Omega) \right\}.$$
Note that because of the standard measure theory convention $\frac{0}{0} = 0$, we have that
$$\sup_{A \in A} \frac{\mu(A)}{Q(A)} = \sup_{\mu(A) > 0} \frac{\mu(A)}{Q(A)} = 1 / \inf_{A \in A, \mu(A) > 0} \frac{Q(A)}{\mu(A)} = e^{D_{\infty}(\mu \parallel Q)}.$$
The claimed identity easily follows.

4.3 Precision-recall vs Lorenz curves

In essence, PR-curves and Lorenz curves are two ways of exposing the couples $(P(A^{Q/P}_\lambda), Q(A^{Q/P}_\lambda))$. Yet, the following questions are not trivial. Given the PR-curve of $P$ and $Q$, can we compute their Lorenz curve? Reciprocally, can we compute the PR-curve from the Lorenz one? If one had a more complete representation such as $(\lambda, P(A^{Q/P}_\lambda), Q(A^{Q/P}_\lambda))$, then one could easily compute both the PR-curve and the Lorenz curve, but in each representation, at least one datum is not explicitly known:

1. In the Lorenz curve, $\lambda$ is not readily available, but we will see that it can be recovered as the derivative of the Lorenz curve.
2. In the PR-curve, $\lambda$ can be easily computed as the ratio $\frac{\alpha}{\beta}$ but the values of $P(A^{Q/P}_\lambda)$ and $Q(A^{Q/P}_\lambda)$ are mingled within $\alpha$ so that one needs to untangle them before recovering the Lorenz curve.

We will use the following result from [13, Theorem 5]:

Theorem 4. Let $P, Q$ two distributions from $M_p(\Omega)$. Then $\forall \lambda \in R^+$,
$$\alpha_\lambda = \min_{A \in A} \lambda (1 - P(A)) + Q(A)$$
From this theorem the following corollary is easily obtained:

Corollary 1. Let $P, Q$ two distributions from $M_p(\Omega)$. Then $\forall \lambda \in R^+$,
$$\alpha_\lambda = \min_{0 \leq f \leq 1} \lambda (1 - \int f dP) + \int f dQ$$
where the functions $f$ are measurable.
\( \lambda_t \) where we have used that if according to Theorem 4, this inequality holds for indicator functions. Besides it is stable under convex combinations and \( L^\infty \) limits. As a result, we can extend the inequality first to convex combinations of indicators which is to say to all simple functions ranging in \([0, 1]\). Then, using standard density results we can further extend it to \( L^\infty \) functions ranging in \([0, 1]\).

From Corollary 4 one can draw the following link between the PR-curve and the Lorenz curve.

**Theorem 5.** Let \( P \) and \( Q \) two distributions. Let \( \lambda \in \mathbb{R}^+ \). Consider the Lorenz Curve \( F \) defined in Eq. (10), then,

\[
F^*(\lambda) = \lambda - \alpha_\lambda
\]

where \( F^*(\lambda) = \sup_{t \in [0, 1]} \lambda t - F(t) \) is the Legendre transform of \( F \).

**Proof.** Let \( \lambda \geq 0 \). Let us show that \( F^*(\lambda) = \lambda - \alpha_\lambda \). Indeed, \( \forall t \in [0, 1] \)

\[
\lambda t - F(t) = \lambda t - \inf_{0 \leq f \leq 1} \int f dP = \sup_{0 \leq f \leq 1} \lambda t - \int f dQ
\]

\[
\leq \sup_{0 \leq f \leq 1} \lambda \int f dP - \int f dQ
\]

\[
\leq \sup_{0 \leq f \leq 1} \lambda \int f dP - \int f dQ
\]

\[
= \lambda - \alpha_\lambda \quad \text{(thanks to Corollary 4)}
\]

Which shows that \( \lambda - \alpha_\lambda \geq \sup_{t \in [0, 1]} \lambda t - F(t) \). Besides, letting \( t_\lambda := P(A_\lambda^{Q/P}) \)

\[
\lambda - \alpha_\lambda = \lambda - (\lambda(1 - P(A_\lambda^{Q/P}) + Q(A_\lambda^{Q/P})) = \lambda P(A_\lambda^{Q/P}) - Q(A_\lambda^{Q/P}) = \lambda t_\lambda - F(t_\lambda)
\]

where we have used that if \( t_\lambda = P(A_\lambda^{Q/P}) \) then \( F(t_\lambda) = Q(A_\lambda^{Q/P}) \) (a result induced by the standard Neyman-Pearson lemma). Therefore, \( \lambda - \alpha_\lambda = \sup_{t \in [0, 1]} \lambda t - F(t) = F^*(\lambda) \).

**Remark 3.** Theorem 5 brings many valuable prospects concerning the link between the PR and Lorenz curves.

1. First, since the Legendre transform is a one-to-one involution, the PR and Lorenz curves are theoretically equivalent.

2. Besides, letting \( t_\lambda := P(A_\lambda^{Q/P}) \) and relying on the Fenchel identity, one gets that \( \lambda \in \partial F(t_\lambda) \), which theoretically provides a means to extract the missing datum as soon as one is capable of computing the subdifferential of the Lorenz curve.

3. More concretely, the theorem provides us a practical way to compute \( \alpha_\lambda \) from the Lorenz curve. Indeed, given \( \lambda \), \( \alpha_\lambda \) can be computed by solving the following 1D convex problem:

\[
\alpha_\lambda = \lambda - F^*(\lambda) = \min_{t \in [0, 1]} F(t) + \lambda(1 - t)
\]

One can do so efficiently thanks to the bisection method if the subdifferential of \( F \) is available or resort to derivative-free algorithms such as the Golden Section Search method otherwise. Then \( \beta_\lambda \) is obtained as \( \tfrac{\alpha_\lambda}{\lambda} \).

4. In the other way around, given \( t \in [0, 1] \), one can solve for \( F(t) \) by considering the following 1D concave problem:

\[
F(t) = F^{**}(t) = \sup_{\lambda \in \mathbb{R}^+} \lambda t - F^*(\lambda) = \sup_{\lambda \in \mathbb{R}^+} \alpha_\lambda + \lambda(t - 1)
\]

At first glance, simple functions ranging in \([0, 1]\) take the form of sub-convex combinations of indicators, but one can leverage \( 1_0 = 0 \) to express them as convex combinations.
5 Conclusion

In this work, we have studied the interconnections among several trade-off curves designed to evaluate the similarity between two probability distributions, namely precision-recall curves, divergence frontiers, ROC and Lorenz curves. If one connection was known to the authors of divergence frontiers, others appear to have eluded even the authors of the implied notions. This is particularly striking for Lorenz and ROC curves which differ by mere symmetries. The interrelation between precision-recall and Lorenz curves is less direct, as it involves convex duality. That being said, it remains that the two notions are theoretically equivalent, and can be computed in practice from one another. We hope that the exposed link will foster new research avenues for evaluation curves. To begin with, while the theoretical equivalence of Lorenz and PR curves has been demonstrated, the question of their empirical estimation has yet to be examined. For instance, investigations on potentially consistent estimators need consideration, especially in the non parametric case. In particular, exposing rates of convergence of the estimators should be a worthwhile endeavor. Similar analysis has already been carried out for scalar metrics [10, 12]. We foresee that exploring links with divergences or integral probability metrics shall help in pursuing this undertaking for PR and Lorenz curves. In any event, evaluation curve estimators with good properties could prove very handy in the design of novel loss functions, hence allowing for an effective control of both fidelity and variability during the training of a generative model.

A Proof of proposition $\textbf{I}$

Proof. The proof is technical and can be skipped on first reading. The second result is a simple corollary of the first since when $\mu$ is not absolutely continuous w.r.t $\nu$ then the identity is trivial: $\infty = \infty$. Let us demonstrate the first point, that is to say when $\mu \ll \nu$. We shall proceed by proving two opposite inequalities, starting from the following.

$$\sup_{A \in \mathcal{A}} \frac{\mu(A)}{\nu(A)} \leq \text{ess sup}_{\nu} \frac{d\mu}{d\nu}.$$ 

Indeed, let $A \in \mathcal{A}$, then,

$$\frac{\mu(A)}{\nu(A)} = \int_A \frac{d\mu}{d\nu} d\nu = \int_A \frac{1}{\nu(A)} \frac{d\mu}{d\nu} d\nu \leq \text{ess sup}_{\nu} \frac{d\mu}{d\nu} \leq \text{ess sup}_{\nu} \frac{d\mu}{d\nu}.$$ 

Note that the essential sup is w.r.t $\nu$ instead of $\mu$. Let us then show that $\text{ess sup}_{\nu} \frac{d\mu}{d\nu} \leq \text{ess sup}_{\mu} \frac{d\mu}{d\nu}$. To do so we need to show that any upper-bound $M$ of $\frac{d\mu}{d\nu}$ $\mu$-a.e. is also an upper-bound $\nu$-a.e.. Let $M \geq \frac{d\mu}{d\nu}$ $\mu$-a.e. be such an upperbound and let $N$ the associated $\mu$-nullset (where $M$ may be lesser than $\frac{d\mu}{d\nu}$). If $\nu(N) = 0$ it settles it (as $M$ is henceforth an upper-bound $\nu$-a.e.). Otherwise, $N$ is such that $\mu(N) = 0$ but $\nu(N) > 0$ and then necessarily $\frac{d\mu}{d\nu} 1_N = 0$ $\nu$-a.e. (or else it would contradict $\mu(N) = 0$). As a result,

$$\frac{d\mu}{d\nu} \leadsto \frac{d\mu}{d\nu} 1_{\Omega \setminus N} \leq M 1_{\Omega \setminus N} \leq M \nu$$-a.e.

Which again settles the fact that $M$ is also an upperbound $\nu$-a.e.. Therefore we obtain that $\text{ess sup}_{\nu} \frac{d\mu}{d\nu} \leq \text{ess sup}_{\mu} \frac{d\mu}{d\nu}$ and $\frac{\mu(A)}{\nu(A)} \leq \text{ess sup}_{\mu} \frac{d\mu}{d\nu}$.

For the reverse inequality, we only need to show that $\frac{d\mu}{d\nu} \leq \sup_{A \in \mathcal{A}} \frac{\mu(A)}{\nu(A)} \mu$-a.e.. Indeed, let $C := \sup_{A \in \mathcal{A}} \frac{\mu(A)}{\nu(A)}$. If $C = \infty$, the inequality is trivial. Let us then suppose that $C < \infty$. Let $E := \{ \omega \in \Omega / \frac{d\mu}{d\nu} \geq C \}$ and let us show that $\mu(E) = 0$. One can rewrite $E = \cup_{n \in \mathbb{N}} E_n$, with $E_n := \{ \omega \in \Omega / \frac{d\mu}{d\nu} \geq C + \frac{1}{n} \}$, and it suffices to show that $\mu(E_n) = 0$. We have that,

$$C \nu(E_n) \geq \mu(E_n) = \int_{E_n} \frac{d\mu}{d\nu} d\nu \geq \left( C + \frac{1}{n} \right) \int_{E_n} d\nu \geq (C + \frac{1}{n}) \nu(E_n)$$

For this not to be absurd, it is necessary that $\nu(E_n) = 0$ and hence $\mu(E_n) = 0$ as well.  \( \square \)
Inputs: Likelihood ratio function: \( \frac{dQ}{dP}(\omega) \),
Sampling routines: sampleFromP() sampleFromQ(),
Likelihood ratio: \( \lambda \in \mathbb{R}^{+} \),
Number of random samples: \( K_{MC} \)

Output: \( (\lambda, P(A_{\lambda}^{Q/P}), Q(A_{\lambda}^{Q/P})) \)

Algorithm estimateLorenzCurve\((p, q, \lambda)\)
1. \( t_\lambda = \text{MCProbaOfLikelihoodRatioSet}(\text{sampleFromP}) \)
2. \( F_{t_\lambda} = \text{MCProbaOfLikelihoodRatioSet}(\text{sampleFromQ}) \)
3. return \( (\lambda, t_{\lambda}, F_{t_{\lambda}}) \)

Procedure MCProbaOfLikelihoodRatioSet\((\text{sampleFromX})\)
1. \( S = 0 \)
2. for \( k = 0; k < K_{MC} \) do
3. \( \omega_k = \text{sampleFromX}() \)
4. \( S = S + \mathbb{1}_{\frac{dQ}{dP}(\omega_k) \leq \lambda} \)
end
5. return \( S_{K_{MC}} \)

Algorithm 1: Monte Carlo computation of the complete representation \( (\lambda, P(A_{\lambda}^{Q/P}), Q(A_{\lambda}^{Q/P})) \)

B Several study cases

This appendix is meant to help the practitioner of Generative models harness the notions involved in our submission. In particular, we propose to elucidate how typical distribution disparities translate in the precision-recall and the Lorenz curve. We are particularly interested in the following aspects:

- **mode dropping**, where part of the distribution \( P \) is entirely missing from the distribution \( Q \),
- **mode invention**, which is the “transposed” case: part of the distribution \( Q \) is missing in \( P \),
- **mode reweighting**, where part of the mass of \( P \) is balanced differently but still present in \( Q \).

In order to illustrate the aforementioned failures and their unveiling under both curve modalities, we consider several configurations where \( P \) and \( Q \) are one-dimensional Gaussian mixtures. We shall present and comment on the different resulting curves in a second stage. But for the sake of completeness, let us first explain how we estimate both curves.

We present here a simple Monte Carlo algorithm that allows to do so for any couple of distributions provided that:

1. One knows how to sample from \( P \) and \( Q \)
2. One can compute likelihood ratios \( \frac{dQ}{dP}(\omega) \)

Please note that in the context of Generative models the first condition is generally true (\( P \) is actually known through a database of samples, and \( Q \) is known through the generative model). However, the second is not realistic, since likelihood estimation remains as yet an open problem\(^8\).

That being said, the purpose of our algorithm is merely to provide a practical way to illustrate the shape of the Lorenz and precision-recall curves in a simple context.

Algorithm 1 presents the Monte Carlo estimation of a complete representation \( (\lambda, P(A_{\lambda}^{Q/P}), Q(A_{\lambda}^{Q/P})) \) from which one can extract both the precision-recall and Lorenz curves. With regards to the latter, one merely needs to drop the \( \lambda \) component and consider the parametrized curve \( \{t_\lambda, F(t_\lambda)\} := (P(A_{\lambda}^{Q/P}), Q(A_{\lambda}^{Q/P})) \). On the other hand, to compute the precision-recall curve, one may use Theorem 2 which implies that

\[
\forall \lambda \in \mathbb{R}^{+}, \alpha_\lambda = \lambda (1 - P(A_{\lambda}^{Q/P})) + Q(A_{\lambda}^{Q/P})
\]

\(^8\) Although one could always resort to the common practice of embedding the domain \( \Omega \) into a feature space wherein simple parametric models can decently fit both distributions.
In the following sections, we scrutinize different critical cases (namely mode dropping, mode invention, and reweighting) for which we estimate the corresponding precision-recall and Lorenz curves using Algorithm 1.

**B.1 Mode dropping**

Mode dropping corresponds to a situation where part of the mass of $P$ is entirely missing from $Q$. To create such a situation, we simply consider a mixture of Gaussians with two components for $P$ and a single of those components in $Q$. Note that the common component corresponds to a weighing coefficient of $0.3$ in $P$. The probability density functions of $P$ and $Q$ are depicted at the top of Figure 3. On the bottom row, one can see how mode reweighting manifests itself when considered under the prism of precision-recall curve (left) and of Lorenz curve (right).

In the precision-recall curve, mode dropping is evidenced by the fact that recall remains constant to a value $\beta = 0.3$ as precision evolves from a perfect value $\alpha = 1$ to the worst precision of $0$. Since this situation is pure mode dropping, this transition is sharp (vertical line). Similar observations hold for the Lorenz curve, where the dropped mass $(0.7)$ manifests itself as the horizontal location $t$ where the Lorenz curve $F(t)$ departs from 0. The fact that only pure mode dropping arises is also visible in this curve. It shows up in the form of a perfectly linear transition from $(0.7, 0)$ to $(1, 1)$. Note that the slope of this linear transition $\frac{1}{0.3}$ corresponds exactly the value of $\lambda$ where the sharp transition happens in the precision-recall. This fact is indicative of the duality between the two curves.

![Figure 3: PR and Lorenz Curves in case of mode dropping.](image-url)
B.2 Mode invention

Mode invention is a transposed configuration compared to mode dropping, where the role of $P$ and $Q$ are reversed. We consider therefore a situation where this time $P$ has a single Gaussian component. This component is shared with $Q$ that contains also an additional component. This time the common mass was set to 0.4. The probability density functions, and both precision-recall and Lorenz curves are displayed in Figure 4.

The realization of mode invention is similar to mode dropping with few distinctions. Indeed, in the precision curve, the sharp transition is now horizontal, and occurs for a value of the precision $\alpha = 0.4$ (again the common mass). Besides, in the Lorenz curve the linear transition arises between $(0, 0)$ and $(1, 0.4)$. Again the slope of the linear transition corresponds to the likelihood ratio $\lambda = 0.4$ where the sharp precision-recall transition.

![Figure 4: PR and Lorenz Curves in case of mode invention.](image-url)
B.3 Mode reweighting

To simulate mode reweighting, we make use of two mechanisms available with Gaussian mixtures, namely we use slightly different mixing factors for $P$ and $Q$ as well as small discrepancies in the standard deviations. On the contrary, we kept the Gaussian means identical in $P$ and $Q$. All the densities and curves of this scenario can be found in Figure 5.

First, considering the precision-recall curve, since all modes of $P$ are present in $Q$, one see that a perfect recall of $\beta = 1$ is reached by an almost vertical transition while trading off some precision. Symmetrically, since no mode of $Q$ is missing in $P$, perfect precision $\alpha = 1$ is achieved through a mostly horizontal shift as recall vanishes. Yet, since modes are weighed and scaled differently, the precision-recall trade-off departs from the perfect matching point $(1, 1)$.

Similarly, looking at the Lorenz curve, the absence of mode dropping is confirmed by the fact that $F(t)$ becomes positive right at the origin. The same notion is true for the lack of mode invention, which is testified by the fact that $\lim_{t\to1}F(t) = 1$. Now, the fact that different reweighting happens is revealed by the non linearity of $F(t)$, in other words the slope of $F(t)$ varies and exposes the achievable likelihood ratios between $Q$ and $P$.

![Figure 5: PR and Lorenz Curves in case of mode reweighting.](image)
B.4 Mix of all failures

As a last study case, we consider a combination of all the previous distribution discrepancies between $P$ and $Q$. Figure 6 gathers the densities and curves in this setting.

Again one can observe that the two curves convey the same information but in a different way. First, half of the mass of $Q$ is pure mode invention, which translates into the fact that both curves are upper bounded by 0.5. Second, as $Q$ is missing nearly 60\% of the mass from $P$ (mode dropping), the recall maxes out around 0.4 and the Lorenz curve takes off approximately at 0.6.

In between, due to the difference in weights of the two common modes, the curves are mainly composed of three parts, the distinction of which is more noticeable on the precision-recall curve. Starting from an almost flat part for large values of $\lambda$ indicating mode invention, the precision-curve exhibits a slope around $\lambda = 1$ showing the trade-off between precision and recall in the region of the common mass due to reweighting before dropping almost vertically when $\lambda$ is close to 0, accounting to mode dropping. These three parts corresponds roughly to three different slopes in the Lorenz curve.

![PR and Lorenz Curves in case of a combination of several mode failures.](image)

Figure 6: PR and Lorenz Curves in case of a combination of several mode failures.

References

[1] A. Brock, J. Donahue, and K. Simonyan, “Large scale gan training for high fidelity natural image synthesis,” *arXiv preprint arXiv:1809.11096*, 2018.

[2] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.
[3] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, “Cnn-generated images are surprisingly easy to spot... for now,” arXiv preprint arXiv:1912.11035, 2019.

[4] R. Webster, J. Rabin, L. Simon, and F. Jurie, “Detecting overfitting of deep generative networks via latent recovery,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[5] A. Borji, “Pros and cons of gan evaluation measures,” Computer Vision and Image Understanding, vol. 179, pp. 41–65, 2019.

[6] 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, 2017, pp. 6626–6637.

[7] M. S. M. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, and S. Gelly, “Assessing generative models via precision and recall,” in Advances in Neural Information Processing Systems 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 5234–5243.

[8] D. J. Im, H. Ma, G. Taylor, and K. Branson, “Quantitatively evaluating gans with divergences proposed for training,” arXiv preprint arXiv:1803.01045, 2018.

[9] M. F. Naeem, S. J. Oh, Y. Uh, Y. Choi, and J. Yoo, “Reliable fidelity and diversity metrics for generative models,” arXiv preprint arXiv:2002.09797, 2020.

[10] P. Rubenstein, O. Bousquet, J. Djolonga, C. Riquelme, and I. O. Tolstikhin, “Practical and consistent estimation of f-divergences,” in Advances in Neural Information Processing Systems, 2019, pp. 4072–4082.

[11] A. Müller, “Integral probability metrics and their generating classes of functions,” Advances in Applied Probability, vol. 29, no. 2, pp. 429–443, 1997.

[12] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. R. Lanckriet, “On integral probability metrics,\phi-divergences and binary classification,” arXiv preprint arXiv:0901.2698, 2009.

[13] L. Simon, R. Webster, and J. Rabin, “Revisiting precision recall definition for generative modeling,” in Proceedings of the 36th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. Long Beach, California, USA: PMLR, 09–15 Jun 2019, pp. 5799–5808.

[14] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, “Improved precision and recall metric for assessing generative models,” in Advances in Neural Information Processing Systems, 2019, pp. 3929–3938.

[15] J. Djolonga, M. Lucic, M. Cuturi, O. Bachem, O. Bousquet, and S. Gelly, “Evaluating generative models using divergence frontiers,” arXiv preprint arXiv:1905.10768, 2019.

[16] Z. Lin, A. Khetan, G. Fanti, and S. Oh, “Pacgan: The power of two samples in generative adversarial networks,” in Advances in Neural Information Processing Systems, 2018, pp. 1498–1507.

[17] Z. Lin, A. Khetan, G. C. Fanti, and S. Oh, “Pacgan: The power of two samples in generative adversarial networks,” CoRR, vol. abs/1712.04086, 2017. [Online]. Available: http://arxiv.org/abs/1712.04086

[18] P. Harremoës, “A new look on majorization,” Proceedings ISITA 2004, Parma, Italy, pp. 1422–1425, 2004.

[19] T. van Erven and P. Harremoës, “Rényi divergence and majorization,” in 2010 IEEE International Symposium on Information Theory. IEEE, 2010, pp. 1335–1339.

[20] B. Piccoli, F. Rossi, and M. Tournus, “A wasserstein norm for signed measures, with application to nonlocal transport equation with source term,” arXiv preprint arXiv:1910.05105, 2019.

[21] M. O. Lorenz, “Methods of measuring the concentration of wealth,” Publications of the American statistical association, vol. 9, no. 70, pp. 209–219, 1905.