

Research Paper

A tale of two curves: Exploring the unexpected relationship between the Lorenz curve and the ROC curve

ZAHRA BEHDANI¹, VAHIDEH AHRARI^{*,2}, GHOLAM REZA MOHTASHAMI BORZADARAN³

¹DEPARTMENT OF MATHEMATICS AND STATISTICS, FACULTY OF ENERGY AND DATA SCIENCES, BEHBAHAN KHATAM ALANBIA UNIVERSITY OF TECHNOLOGY, BEHBAHAN, IRAN

²DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF MATHEMATICAL SCIENCES, SHAHREKORD UNIVERSITY, SHAHREKORD, IRAN

³DEPARTMENT OF STATISTICS, FACULTY OF MATHEMATICAL SCIENCES, FERDOWSI UNIVERSITY OF MASHHAD, IRAN

Received: May 05, 2026/ Revised: September 06, 2026/ Accepted: September 18, 2026

Abstract: This paper provides a conceptual and methodological synthesis of the relationship between two tools that originate from different disciplines: the Lorenz curve, widely used to study economic inequality, and the receiver operating characteristic curve, widely used to evaluate binary classification and predictive discrimination. Building on established results linking the Gini coefficient to the area under the receiver operating characteristic curve, the study examines their relationship from a broader ranking-based and distribution-based perspective, clarifying their mathematical connections, distributional foundations, and distinct interpretations through analytical examples, Monte Carlo simulations, and applied evaluations under different sample sizes and class imbalance conditions. By clarifying these connections and examining their empirical behavior beyond idealized settings, the study provides a cross-disciplinary framework for interpreting ranking-based measures in both inequality measurement and predictive classification.

Keywords: Gini coefficient; Lorenz curve; Predictive classification; Ranking measures; Receiver operating characteristic curve.

Mathematics Subject Classification (2010): 62H30, 62G10, 91B82.

*Corresponding author: vahideh.ahrari@sku.ac.ir

1 Introduction

Receiver operating characteristic (ROC) curve analysis is a widely used approach for evaluating the discriminatory ability of binary classification and diagnostic systems. The ROC curve is obtained by plotting the true positive rate (TPR) against the false positive rate (FPR) across decision thresholds, providing a threshold-independent measure of the ability to distinguish between two populations. ROC analysis has been extensively applied in medicine, engineering, machine learning, data mining, and other predictive domains. Originating from signal detection theory, ROC methodology was later adopted in psychological research and diagnostic medicine, particularly in radiology and diagnostic test evaluation (Goodenough et al., 1974; Hanley and McNeil, 1982, 1983; Erdreich, 1981). Its importance has further increased with the development of modern predictive models and classification algorithms.

The area under the ROC curve (AUC), which is discussed in ROC analysis by Hanley and McNeil (1982) and Fawcett (2006), is one of the most commonly used summary measures for evaluating discrimination. It provides a global and threshold-independent measure and can be interpreted as the probability that a randomly selected positive observation receives a higher score than a randomly selected negative observation. Higher AUC values indicate stronger discriminatory performance, with $AUC = 1$ representing perfect discrimination. Therefore, AUC has become a standard measure for comparing diagnostic procedures and binary classifiers.

The Lorenz curve, introduced by Lorenz (1905), provides a fundamental framework for analyzing economic inequality and concentration through the cumulative distribution function (CDF) of a non-negative random variable. Under the assumptions of a non-negative support and finite positive mean, the Lorenz curve and the Gini coefficient have become standard tools for measuring income and wealth inequality since the contributions of Lorenz (1905) and Gini (1914). The Gini coefficient summarizes the deviation of the Lorenz curve from perfect equality and quantifies the concentration of a non-negative quantity within a population.

The study of inequality measurement has resulted in various extensions of Lorenz curves and alternative concentration indices. Theoretical properties of Lorenz curves, their moments, and related inequality measures have been extensively investigated. As emphasized by Arnold and Sarabia (2018), distinguishing between population-level theoretical results and sample-based statistical quantities is essential, particularly when examining relationships between inequality measures and statistical performance measures in finite samples.

Although ROC and Lorenz analyses originate from different fields, their mathematical structures share common elements, including CDFs, ranking, and comparisons between populations. These connections have motivated studies linking Lorenz representations, ROC curves, concentration measures, and related indices. Kämpke and Radermacher (2015) examined Lorenz-based inequality measures, while Figini et al. (2012) studied ROC ordering and stochastic dominance. More directly, Irwin and Hautus (2015) demonstrated that under specific distributional assumptions, lognormal Lorenz curves can correspond to normal ROC representations, providing an important link between inequality and discrimination frameworks.

The relationship between Gini-type measures and ROC analysis has been studied from a concentration perspective. Schechtman and Schechtman (2019) examined

the connection between Gini terminology and ROC curves, while Gasparini and Sacchetto (2020, 2023) developed ROC-related concentration functions and explored their link with diagnostic discrimination. These studies highlight mathematical connections among Lorenz-type representations, concentration measures, ROC curves, and Gini-related indices.

Despite these established connections, the literature remains divided across disciplinary perspectives. Lorenz curves and Gini indices are mainly studied in inequality and concentration analysis, whereas ROC curves and AUC focus on diagnostic accuracy and predictive discrimination. A unified framework that integrates these perspectives while distinguishing mathematical relationships from substantive interpretations can provide a valuable cross-disciplinary perspective.

An important distinction should be made between the inequality Gini coefficient and the Gini discrimination index. In inequality analysis, the Gini coefficient is associated with the Lorenz curve, whereas in classification and ranking analysis the quantity

$$G_{\text{disc}} = 2 \text{AUC} - 1, \quad (1)$$

represents a Gini discrimination or ranking index. Although both measures share mathematical and geometric connections, they have different substantive interpretations: the former describes distributional inequality, while the latter quantifies the ranking ability of a score to distinguish between two classes.

The identity given in (1) is an established result and is not considered a new mathematical contribution of this study. Instead, this paper examines how Lorenz-type curves, concentration measures, ROC curves, AUC, and Gini-related indices can be interpreted within a unified ranking- and distribution-based framework, while considering related threshold-dependent measures such as Cohen (1960).

The paper complements the conceptual analysis with analytical examples, simulation experiments, and applied illustrations. The empirical study evaluates the finite-sample behavior of AUC and its Gini discrimination transformation under varying distributional conditions, sample sizes, and class imbalance, focusing on bias, variability, and confidence-interval coverage.

Accordingly, the manuscript is organized as follows. The next sections introduce the Lorenz curve and ROC framework, including their main properties and associated measures. Subsequent sections develop the mathematical and conceptual connections among ROC representations, Lorenz-type curves, Gini-related indices, and concentration measures, followed by analytical examples, simulations, and applied illustrations. The final section summarizes the main findings and concluding remarks.

The relationship between AUC and the Gini discrimination index, given in (1), is an established result and should be distinguished from the economic Gini coefficient used in inequality measurement. This study does not introduce this transformation as a new theoretical result instead, it provides an integrated analytical perspective on the connections among Lorenz-type representations, inequality and discrimination Gini measures, ROC curves, AUC, and concentration functions.

The primary contribution of this study is integrative, combining perspectives from inequality measurement, concentration analysis, diagnostic testing, and classification within a common framework based on CDFs, ordering, and ranking. It synthesizes established results from Lorenz and Gini analysis (Lorenz, 1905; Gini, 1914; Arnold

and Sarabia, 2018), ROC analysis (Hanley and McNeil, 1982; Fawcett, 2006), Lorenz–ROC relationships (Irwin and Hautus, 2015), Gini–ROC connections (Schechtman and Schechtman, 2019), and ROC-based concentration functions (Gasparini and Sacchetto, 2020, 2023). The aim is not to replace these contributions, but to clarify their relationships and the conditions under which their mathematical similarities admit meaningful interpretations.

A second contribution is the distinction between mathematical equivalence and substantive interpretation. The established transformation in (1) does not imply that economic inequality and predictive discrimination represent the same phenomenon. The economic Gini coefficient measures inequality or concentration of a non-negative distribution, whereas AUC and G_{disc} quantify the ranking ability of a score to distinguish between two populations. This study therefore distinguishes exact identities from interpretive similarities and clarifies the conditions under which Lorenz-type and ROC-type representations can be meaningfully compared.

A third contribution is empirical. The study examines the finite-sample behavior of AUC and its Gini discrimination transformation through Monte Carlo simulations under varying sample sizes, class imbalance levels, and distributional settings. The analysis evaluates bias, sampling variability, and confidence-interval coverage, while characterizing how uncertainty is transferred between AUC and G_{disc} through their linear relationship.

Finally, the study provides analytical and applied illustrations to clarify the practical interpretation of the developed relationships. The comparison with Cohen’s kappa highlights the distinction between threshold-independent discrimination and threshold-dependent agreement. Table 1 summarizes representative contributions and positions this study as an integrative synthesis of established results, combined with empirical robustness analysis, rather than a proposal of a new Gini-AUC identity.

2 The Lorenz curve: A measure of inequality

Introduced by Max Lorenz in 1905, the Lorenz curve is a fundamental graphical representation of the distribution of a non-negative quantity, such as income or wealth, within a population (Lorenz, 1905). In its standard form, the horizontal axis represents the cumulative population share ordered from the lowest to the highest values, while the vertical axis represents the corresponding cumulative share of the total quantity. The curve starts at $(0, 0)$ and ends at $(1, 1)$, with the 45° line representing perfect equality, where each population proportion receives an equal share of the total amount (Lorenz, 1905).

For a non-negative random variable X with distribution function F and finite positive mean μ , the population Lorenz curve is defined by

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(u) du, \quad 0 \leq p \leq 1,$$

where $F^{-1}(u)$ denotes the quantile function of X . The Lorenz curve satisfies $L(0) = 0$ and $L(1) = 1$. Under standard assumptions, $L(p)$ is non-decreasing and convex; however, differentiability is not required everywhere because some distributions may

Table 1: Representative literature on Lorenz curves, Gini-related measures, ROC curves, AUC, and concentration functions.

Study	Perspective and contribution	Relevance to the present study
Lorenz (1905)	<i>Lorenz curve and concentration.</i> Introduced the Lorenz curve for representing wealth concentration and distribution.	Foundation for the inequality and concentration perspective.
Gini (1914)	<i>Economic inequality measurement.</i> Developed the Gini coefficient as a scalar measure of distributional inequality.	Foundation for the Lorenz-based inequality measure.
Arnold and Sarabia (2018)	<i>Lorenz ordering and majorization.</i> Developed the theoretical treatment of Lorenz ordering and majorization in inequality and distributional analysis.	Supports the distinction between population-level Lorenz concepts and sample-based quantities.
Hanley and McNeil (1982, 1983)	<i>ROC curves and AUC.</i> Established fundamental interpretations and statistical methods for AUC in diagnostic discrimination.	Statistical foundation for interpreting AUC as a discrimination measure.
Fawcett (2006)	<i>ROC analysis and classification.</i> Provided a systematic treatment of ROC curves, AUC, thresholds, and classification performance.	Provides the classification and machine-learning perspective.
Figini et al. (2012)	<i>Stochastic comparison of ROC curves.</i> Studied ordering and stochastic comparison of ROC curves, including intersecting curves.	Background for distributional ordering and ROC comparisons.
Irwin and Hautus (2015)	<i>Lorenz-ROC relationship.</i> Established a mathematical relationship between lognormal Lorenz curves and normal ROC curves under specific distributional assumptions.	Key example of a distribution-specific Lorenz-ROC connection.
Schechtman and Schechtman (2019)	<i>Gini terminology and ROC curves.</i> Examined Gini-related terminology, ROC curves, and concentration-based interpretations of ROC analysis.	Supports distinction between Gini-related discrimination concepts and ROC interpretations.
Gasparini and Sacchetto (2020)	<i>Concentration functions and ROC.</i> Defined a concentration function for ROC analysis and examined its theoretical properties.	Connects ROC analysis with concentration concepts.
Gasparini and Sacchetto (2023)	<i>Concentration and ROC curves.</i> Further examined the relationship between concentration functions and ROC curves.	Extends the concentration-based theoretical perspective.
Present study	<i>Integrated analytical and methodological synthesis.</i> Brings together Lorenz-type representations, the economic Gini coefficient, Gini-related discrimination measures, ROC curves, AUC, concentration functions, and Cohen's kappa through analytical, simulation-based, and applied investigations.	Provides a unified framework, separates mathematical identities from interpretations, and evaluates robustness.

yield non-differentiable quantile functions or Lorenz curves. When differentiability holds, $L'(p) = \frac{F^{-1}(p)}{\mu}$.

The value $L(p)$ represents the cumulative proportion of the total amount held by the lowest 100 p % of the population. The deviation of the Lorenz curve from the line of perfect equality reflects the degree of inequality: curves closer to this line indicate more equal distributions, whereas larger deviations imply greater inequality. Several related concentration and inequality measures can also be derived from the Lorenz curve.

- **Leimkühler curve:** $K(p) = 1 - L(1 - p)$, $0 \leq p \leq 1$, representing the cumulative share of the upper population when ordered from highest to lowest values.

- **Bonferroni curve:** $B(p) = \frac{L(p)}{p}$, $0 < p \leq 1$, comparing the average amount held by the lowest proportion p with the population mean.

- **Zenga curve:** $Z(p) = 1 - \frac{L(p)}{p} \frac{1-p}{1-L(p)}$, $0 < p < 1$, providing an alternative inequality representation based on lower and upper population segments.

These extensions show that Lorenz-based representations provide alternative views of concentration and inequality. In this study, they are considered primarily as conceptual foundations for comparing Lorenz-type representations, concentration functions, and ROC-based measures.

2.1 Gini coefficient

The Gini coefficient, introduced by Corrado Gini and widely used as a measure of inequality, quantifies the deviation of the Lorenz curve from the line of perfect equality (Lorenz, 1905; Gini, 1914). For a Lorenz curve $L(p)$, the population Gini coefficient is defined as

$$G_{\text{ineq}} = 1 - 2 \int_0^1 L(p) dp.$$

For a non-negative distribution, G_{ineq} ranges from 0 to 1, where 0 represents perfect equality and larger values indicate greater inequality. The coefficient is widely used to compare inequality and concentration across populations and distributions.

2.2 Distinguishing income Gini from Gini discrimination

A central distinction in this study concerns two different quantities that are both referred to as “Gini.” The inequality Gini coefficient, denoted by G_{ineq} , is derived from the Lorenz curve and measures inequality in the distribution of a non-negative variable, whereas the Gini discrimination index, denoted by G_{disc} , is associated with ROC-based ranking performance. Table 2 summarizes the distinction between these measures.

The established transformation between AUC and G_{disc} is given by (1) and does not imply equivalence between economic inequality and predictive discrimination. Although both frameworks involve CDFs, ordering, and concentration-related concepts, they address different quantities and scientific objectives. Therefore, G_{ineq} and G_{disc} are treated as distinct measures throughout this study.

Table 2: Comparison between the inequality Gini coefficient and the Gini discrimination index.

Feature	Inequality Gini coefficient	Gini discrimination index
Primary quantity	Income, wealth, or other non-negative quantity	Predictive or classification score ranking two classes
Main purpose	Measures inequality or concentration	Measures discriminatory ranking ability
Underlying representation	Lorenz curve	ROC curve and ranking comparison
Interpretation	0: perfect equality; larger values: greater inequality	0: random discrimination; 1: perfect discrimination when higher scores indicate the positive class
Substantive meaning	Distributional inequality or concentration	Ability to rank positive observations above negative observations

3 The ROC curve: Evaluating classification performance

The ROC curve is a fundamental tool for evaluating the discriminatory performance of binary classifiers in statistics, diagnostic testing, and machine learning (Metz, 1978; Hanley and McNeil, 1982; Fawcett, 2006). It represents the relationship between TPR and FPR across classification thresholds and provides a threshold-independent assessment of ranking performance. The main quantities defining the ROC curve are:

- **True Positive Rate or Sensitivity:** $\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$.
- **False Positive Rate:** $\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 1 - \text{Specificity}$.

The ROC curve is obtained by plotting TPR against FPR across all possible thresholds and extends from $(0, 0)$ to $(1, 1)$. Curves closer to the upper-left corner indicate stronger discrimination, whereas $\text{TPR} = \text{FPR}$ represents random classification. Since ROC analysis evaluates the ranking of observations across all thresholds rather than a single cutoff, it provides the foundation for the threshold-independent measure AUC.

3.1 Area under the curve

The AUC is a widely used measure of binary discrimination (Hanley and McNeil, 1982; Fawcett, 2006). It can be interpreted as the probability that a randomly selected positive observation receives a higher score than a randomly selected negative observation. Accounting for possible ties,

$$\text{AUC} = P(S_1 > S_0) + \frac{1}{2}P(S_1 = S_0). \quad (2)$$

For continuous score distributions, this reduces to $\text{AUC} = P(S_1 > S_0)$. Equivalently, AUC represents the area under the ROC curve:

$$\text{AUC} = \int_0^1 \text{ROC}(t) dt,$$

where t denotes the FPR. In general, $0 \leq \text{AUC} \leq 1$, with 0.5 indicating random discrimination and 1 representing perfect discrimination. The probabilistic interpretation

in (2) highlights that AUC is fundamentally a ranking-based measure, providing the basis for its connection with Gini discrimination measures.

3.2 On the relationship between Lorenz-type and ROC curves

Although Lorenz and ROC curves may exhibit geometric similarities, they are based on different probabilistic structures and should not generally be considered identical. The Lorenz curve is derived from the quantile function of a non-negative random variable and normalized by its mean, whereas the ROC curve is constructed from class-conditional score distributions and represents the trade-off between TPR and FPR across thresholds. Therefore, relationships such as $L(p) = 1 - \text{ROC}(1 - p)$, are not universal identities.

Under specific distributional assumptions and transformations, however, exact correspondences may exist. For example, Irwin and Hautus (2015) established a relationship between lognormal Lorenz curves and ROC representations based on normal distributions. Such cases demonstrate that common distributional structures can generate related representations, but they should be distinguished from general graphical similarities. Accordingly, this study considers Lorenz–ROC connections as valid analytical correspondences only when the underlying assumptions and transformations are explicitly specified, avoiding the interpretation of graphical resemblance as a general mathematical identity.

4 Mathematical relationship between the ROC curve, Leimkühler curve, and Cohen’s Kappa

The Lorenz and ROC curves originate from different fields and address different objectives: the former describes the distribution and concentration of a non-negative variable, whereas the latter evaluates the ranking ability of binary classifiers. Despite these differences, both can be represented through CDFs and rankings. The established relationship between ROC-based discrimination and the Gini discrimination index is given by (1) and is not presented as a new result of this study. Related connections among ROC curves, concentration measures, and Lorenz-type representations have been studied under specific definitions and distributional assumptions (Irwin and Hautus, 2015; Schechtman and Schechtman, 2019; Gasparini and Sacchetto, 2020, 2023).

This section focuses on clarifying the relationships among Lorenz curves, Leimkühler curves, ROC-based representations, and Cohen’s kappa, while distinguishing mathematical transformations from substantive interpretations. In particular, the inequality Gini coefficient G_{ineq} should be distinguished from the ranking-based Gini discrimination index G_{disc} .

4.1 Distributional perspectives: Lognormal and normal distributions

The connection between Lorenz-type and ROC-type curves can be illustrated under specific distributional assumptions. Consider a positive random variable X following

a lognormal distribution, i.e. $\log(X) \sim \mathcal{N}(\mu, \sigma^2)$. The Lorenz curve of the distribution has the closed-form expression (Kleiber and Kotz, 2003) as

$$L(p) = \Phi(\Phi^{-1}(p) - \sigma), \quad 0 \leq p \leq 1,$$

where Φ denotes the standard normal CDF and $\Phi^{-1}(p)$ is the quantile function of the standard normal distribution. The parameter μ does not affect the Lorenz curve, and inequality is determined by σ . The corresponding Gini coefficient is

$$G_{\text{ineq}} = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1.$$

Thus, larger σ implies greater inequality in the lognormal distribution. For classification settings, let X_1 and X_0 denote positive and negative class scores. The ROC curve is defined as

$$\text{ROC}(t) = 1 - F_1(F_0^{-1}(1 - t)), \quad 0 \leq t \leq 1,$$

where F_1 and F_0 are the CDFs of the positive (event) and the negative (non-event) classes, respectively (Hanley and McNeil, 1982). For continuous scores, AUC represents the ranking probability

$$\text{AUC} = P(X_1 > X_0),$$

and the corresponding Gini discrimination index follows from (1). This relationship reflects a ranking identity and should not be interpreted as an equivalence between income inequality and predictive discrimination.

4.2 The normal equal-variance case

Assume equal-variance normal class-conditional scores such that $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ and $X_0 \sim \mathcal{N}(\mu_0, \sigma^2)$. The ROC curve and AUC are given by

$$\begin{aligned} \text{ROC}(t) &= \Phi\left(\Phi^{-1}(t) + \frac{\mu_1 - \mu_0}{\sigma}\right), \\ \text{AUC} &= \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{2}\sigma}\right). \end{aligned}$$

The corresponding G_{disc} follows from (1). These expressions highlight the distinction between inequality and discrimination: in the lognormal inequality model, larger σ increases G_{ineq} , whereas in the equal-variance normal classification model, larger σ reduces the standardized separation $d' = (\mu_1 - \mu_0)/\sigma$ and decreases AUC and G_{disc} . For example, with $\mu_1 = 1.2$, $\mu_0 = 0$, and $\sigma = 1$,

$$\text{AUC} = \Phi\left(\frac{1.2}{\sqrt{2}}\right) \approx 0.802,$$

giving $G_{\text{disc}} \approx 0.604$. At threshold $c = 0.5$,

$$\text{TPR} = 1 - \Phi\left(\frac{c - \mu_1}{\sigma}\right) \approx 0.758, \quad \text{FPR} = 1 - \Phi\left(\frac{c - \mu_0}{\sigma}\right) \approx 0.309.$$

This example distinguishes threshold-independent ranking measures such as AUC from threshold-dependent operating characteristics such as TPR and FPR.

4.3 Lorenz-type and Leimkühler representations

The Lorenz curve is defined as

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(u) du, \quad 0 \leq p \leq 1,$$

for a nonnegative random variable with finite positive mean μ . It represents the cumulative share of the total quantity held by the bottom proportion p of the population.

The Leimkühler curve (complementary Lorenz curve) is obtained through the transformation

$$K(p) = 1 - L(1 - p).$$

This transformation reflects a complementary cumulative representation and does not imply that Lorenz and ROC curves describe the same statistical object (Aitchison and Brown, 1957).

More generally, ROC-based cumulative representations may take the form

$$C(p) = 1 - \text{ROC}(1 - p),$$

but such representations should not automatically be identified with Lorenz curves. Their equivalence requires additional definitions and distributional assumptions (Irwin and Hautus, 2015; Schechtman and Schechtman, 2019; Gasparini and Sacchetto, 2020, 2023). For example, under the equal-variance normal classification model,

$$1 - \text{ROC}(1 - p) = \Phi \left(\Phi^{-1}(p) - \frac{\mu_1 - \mu_0}{\sigma} \right).$$

Although this expression is mathematically related to ROC analysis, it is not a Lorenz curve of a normal income distribution, since normal variables do not generally satisfy the nonnegative support required for the usual Lorenz framework.

4.4 Lognormal classification scores

Closed-form expressions for AUC can also be obtained when class scores follow lognormal distributions. Suppose $\log(X_j) \sim \mathcal{N}(\mu_j, \sigma_j^2)$, for $j = 1, 2$. Since the logarithm is strictly increasing,

$$P(X_1 > X_0) = P(\log X_1 > \log X_0).$$

For independent class-conditional scores,

$$\log X_1 - \log X_0 \sim \mathcal{N}(\mu_1 - \mu_0, \sigma_1^2 + \sigma_0^2),$$

which gives

$$\text{AUC} = \Phi \left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}} \right).$$

For equal log-scale variances, $\sigma_1 = \sigma_0 = \sigma$, this reduces to

$$\text{AUC} = \Phi \left(\frac{\mu_1 - \mu_0}{\sqrt{2}\sigma} \right).$$

Thus, non-normal scores may still yield analytical AUC expressions under suitable distributional assumptions. In contrast, the Lorenz curve and Gini coefficient of a lognormal income distribution follow different definitions. Although their closed forms may resemble the normal equal-variance AUC expression, the similarity arises from the underlying Gaussian transformation and does not imply equivalence between inequality and discrimination measures.

4.5 ROC and Cohen's Kappa

Cohen's kappa measures agreement between predicted and true categorical labels after accounting for chance agreement (Cohen, 1960; Pontius and Millones, 2011). It is defined as

$$\kappa = \frac{P_o - P_e}{1 - P_e}, \quad (3)$$

where P_o and P_e denote observed and chance-expected agreement, respectively.

Let $\pi = P(Y = 1)$ denote the positive class prevalence. For a classifier with a fixed threshold,

$$P_o = \pi \text{TPR} + (1 - \pi)(1 - \text{FPR}),$$

and with predicted-positive proportion $\hat{\pi} = \pi \text{TPR} + (1 - \pi)\text{FPR}$, the expected agreement is

$$P_e = \pi \hat{\pi} + (1 - \pi)(1 - \hat{\pi}).$$

Thus, kappa depends on the threshold-specific TPR, FPR, and class prevalence, unlike AUC, which evaluates ranking performance independently of a threshold. Therefore, classifiers with identical AUC values may yield different kappa values.

Under the normal equal-variance model, using

$$\text{TPR} = 1 - \Phi\left(\frac{c - \mu_1}{\sigma}\right), \quad \text{FPR} = 1 - \Phi\left(\frac{c - \mu_0}{\sigma}\right),$$

kappa can be computed analytically once the threshold and prevalence are specified.

For $\mu_1 = 1.2$, $\mu_0 = 0$, $\sigma = 1$, $\pi = 0.3$, and $c = 0.5$, the model gives $\text{TPR} \approx 0.758$ and $\text{FPR} \approx 0.309$. Consequently, $P_o \approx 0.711$, $\hat{\pi} \approx 0.444$, and $P_e \approx 0.523$, resulting in $\kappa \approx 0.395$. This example demonstrates that kappa incorporates threshold and prevalence information that is not contained in AUC alone.

4.6 Comparison of distributional and performance measures

Table 3 summarizes the main mathematical objects considered in this section. The distinction between G_{ineq} and G_{disc} is maintained throughout.

4.7 Interpretational distinction

Despite their mathematical connections, the measures considered above quantify different properties:

- **Lorenz Curve:** describes the CDF of a nonnegative variable ranked by its magnitude.

Table 3: Comparison of Lorenz-, ROC-, and agreement-based measures.

Concept	Normal setting	Lognormal setting
Lorenz curve	Defined for nonnegative variables as $L(p) = \mu^{-1} \int_0^p F^{-1}(u) du$; normal variables are not generally suitable for income modeling.	For $\log X \sim N(\mu, \sigma^2)$, $L(p) = \Phi(\Phi^{-1}(p) - \sigma)$.
Income Gini	G_{ineq} is obtained from the Lorenz curve; no general equivalence with AUC exists.	$G_{\text{ineq}} = 2\Phi(\sigma/\sqrt{2}) - 1$.
ROC curve	Equal-variance normal scores yield a closed-form ROC representation.	Obtained from lognormal class-conditional distributions.
AUC	$\text{AUC} = \Phi((\mu_1 - \mu_0)/(\sqrt{2}\sigma))$.	$\text{AUC} = \Phi((\mu_1 - \mu_0)/\sqrt{\sigma_1^2 + \sigma_0^2})$.
Gini discrimination	$G_{\text{disc}} = 2 \text{AUC} - 1$.	The same AUC-based transformation applies.
Leimkühler curve	$K(p) = 1 - L(1 - p)$.	The same transformation applies to lognormal Lorenz curves.
Cohen's Kappa	$\kappa = (P_o - P_e)/(1 - P_e)$; depends on threshold, TPR, FPR, and prevalence.	Same definition with TPR and FPR from lognormal scores.

- **Leimkühler Curve:** provides a complementary transformation of the Lorenz curve and preserves information from the same distribution.
- **ROC Curve and AUC:** characterize classifier ranking performance, with AUC representing the probability-based discrimination between positive and negative classes.
- **Gini Discrimination Index:** defined by (1), measures ranking discrimination rather than economic inequality.
- **Cohen's Kappa:** measures threshold-specific agreement between predictions and observed labels while accounting for chance agreement.

The common mathematical theme among these measures is the use of CDFs and ranking structures, not identical substantive meaning. Lorenz–ROC correspondences require explicit assumptions and transformations, whereas the AUC-based Gini discrimination identity is a general property of the ROC framework and should be distinguished from the inequality Gini coefficient.

4.8 Limitations and caveats

Several caveats should be considered when interpreting the relationships discussed in this study.

1. **Distinct Gini interpretations:** G_{ineq} and G_{disc} are mathematically and substantively different measures. The identity in (1) does not imply equivalence between economic inequality and predictive discrimination.

2. **Distributional assumptions:** Closed-form relationships between Lorenz- and ROC-type representations depend on specific distributional assumptions and should not be generalized beyond those settings.

3. **Threshold dependence of kappa:** Unlike AUC, Cohen’s kappa depends on a selected threshold and class prevalence and cannot generally be derived from AUC alone.

4. **Interpretation of examples:** The numerical examples illustrate mathematical relationships among measures and do not imply causal connections between inequality and classification performance.

5. **Fairness and scope:** The present study does not evaluate algorithmic fairness; such analyses require dedicated criteria such as demographic parity, equalized odds, and calibration (Hardt et al., 2016; Chouldechova, 2017). Its purpose is instead to synthesize established relationships among Lorenz-, concentration-, ROC-, and agreement-based measures under explicit assumptions.

5 Monte Carlo simulation results

To evaluate the finite-sample behavior of the relationship between AUC and the Gini discrimination index, Monte Carlo simulations were performed under controlled data-generating settings. The analysis includes a baseline Gaussian experiment comparing empirical estimates with theoretical values and an extended simulation assessing the effects of sample size, class imbalance, and distributional characteristics.

5.1 Baseline Gaussian simulation

Synthetic scores were generated from $\mathcal{N}(\mu_0 = 0, \sigma = 1)$ for the negative class and $\mathcal{N}(\mu_1 = 2, \sigma = 1)$ for the positive class, with 1,000 observations per class in each iteration.

For continuous scores, AUC was estimated using the pairwise-ranking estimator

$$\widehat{\text{AUC}} = \frac{\sum_{i=1}^{N_+} \sum_{j=1}^{N_-} I(S_+^i > S_-^j)}{N_+ N_-},$$

and the corresponding Gini discrimination estimator was obtained from (1).

For the equal-variance normal model,

$$\text{AUC} = \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{2}\sigma}\right) = \Phi(\sqrt{2}) \approx 0.921,$$

which corresponds to $G_{\text{disc}} \approx 0.843$. Table 4 shows that empirical estimates closely match theoretical values, with deviations attributable to finite-sample Monte Carlo variation rather than differences from the population relationship. The simulation therefore evaluates estimator behavior rather than establishing the already known AUC–Gini transformation. Figure 1 illustrates the empirical distributions of AUC and the Gini discrimination index, with reference lines indicating their theoretical values.

Table 4: Simulation Results (1,000 Iterations).

Metric	Empirical Value	Theoretical Value
AUC	0.922 ± 0.008	0.921
Gini Discrimination	0.843 ± 0.016	0.843

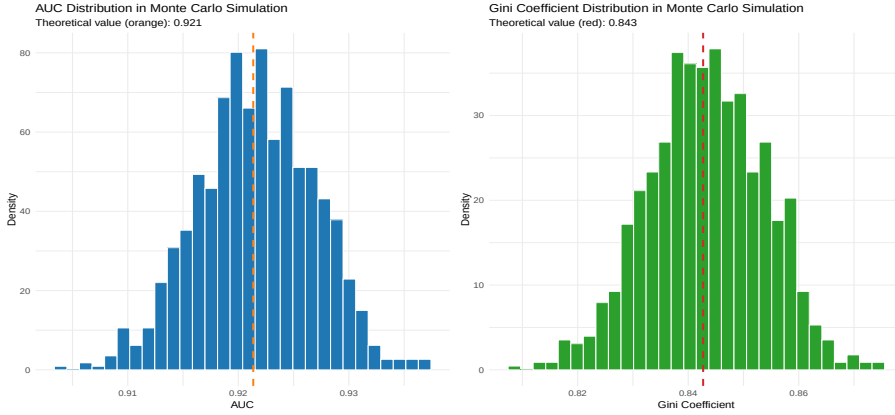


Figure 1: Empirical distributions of (left) AUC and (right) Gini discrimination index from the baseline Monte Carlo simulation. Dashed vertical lines indicate the corresponding theoretical values.

5.2 Expanded finite-sample simulation

The baseline experiment was extended to evaluate the finite-sample behavior of AUC under different sample sizes, class-imbalance levels, and score distributions, including normal, lognormal, exponential, and mixture settings. Minority-class sample sizes of 50, 200, and 1,000 observations were considered.

For each simulation scenario, the empirical bias and variance of the AUC estimator were calculated as

$$\begin{aligned} \text{Bias}(\widehat{\text{AUC}}) &= \frac{1}{R} \sum_{r=1}^R \widehat{\text{AUC}}_r - \text{AUC}_{\text{true}}, \\ \text{Var}(\widehat{\text{AUC}}) &= \frac{1}{R-1} \sum_{r=1}^R \left(\widehat{\text{AUC}}_r - \overline{\widehat{\text{AUC}}} \right)^2. \end{aligned}$$

For nominal 95% confidence intervals, empirical coverage was estimated as

$$\text{Coverage} = \frac{1}{R} \sum_{r=1}^R I(\text{AUC}_{\text{true}} \in CI_r).$$

Because the Gini discrimination index is obtained from AUC through the linear transformation in (1), its bias, variance, and confidence intervals follow directly from the corresponding AUC quantities. Therefore, the simulation results are reported using AUC, while the implications for G_{disc} are determined analytically. Table 5 summarizes the Monte Carlo results across the considered distributional settings, sample sizes, and imbalance ratios.

Table 5: Monte Carlo results for AUC estimation under different sample sizes, class imbalance ratios, and score distributions.

Distribution	$n_{\min}$	Ratio	Bias(AUC)	Var(AUC)	Cov(AUC)
Normal	50	1:1	-0.000296	0.000773	0.914
Normal	50	10:1	-0.000303	0.000394	0.916
Normal	50	100:1	-0.000251	0.000372	0.906
Normal	200	1:1	-0.000160	0.000186	0.930
Normal	1000	1:1	0.000211	0.000032	0.958
Lognormal	50	1:1	0.001154	0.002585	0.942
Lognormal	200	1:1	0.000407	0.000646	0.948
Lognormal	1000	1:1	0.002142	0.000128	0.944
Exponential	50	1:1	-0.003209	0.002932	0.946
Exponential	200	1:1	0.000424	0.000736	0.956
Exponential	1000	1:1	0.000840	0.000140	0.950
Mixture	50	1:1	-0.000967	0.001195	0.920
Mixture	200	1:1	-0.000102	0.000269	0.950
Mixture	1000	1:1	0.000233	0.000054	0.944

5.3 Results and robustness assessment

The results in Table 5 indicate that the AUC estimator exhibits small empirical bias across the considered scenarios. For the normal distribution, the absolute bias remains below 0.001 in all reported settings, while the largest observed deviation occurs for the exponential distribution at $n_{\min} = 50$, with $|\text{Bias}(\widehat{\text{AUC}})| = 0.003209$. These deviations are small relative to the magnitude of AUC.

Estimator variability decreases with increasing sample size. Under the balanced normal setting, the empirical variance decreases from 0.000773 at $n_{\min} = 50$ to 0.000186 and 0.000032 at $n_{\min} = 200$ and 1000, respectively. Similar patterns are observed across the other distributional settings, indicating improved estimation stability with larger sample sizes.

The imbalance scenarios do not indicate a simple monotonic relationship between imbalance ratio and AUC variance. For fixed $n_{\min} = 50$ under the normal distribution, the empirical variances for ratios 1:1, 10:1, and 100:1 are 0.000773, 0.000394, and 0.000372, respectively. Therefore, imbalance effects should be interpreted together with the absolute class sizes rather than through the imbalance ratio alone.

The empirical coverage probabilities of the nominal 95% confidence intervals range from 0.906 to 0.958. Although several values are close to the nominal level, deviations occur in some finite-sample settings, emphasizing the importance of evaluating interval performance in addition to point estimation.

Since G_{disc} is a linear transformation of AUC through (1), its finite-sample bias, variance, and confidence-interval coverage are determined directly by the corresponding AUC results. Therefore, the Gini discrimination results are not repeated in Table 5 to avoid redundant reporting. Figure 2 summarizes the empirical bias and variance patterns across the simulation scenarios. As shown in Figure 2, empirical bias remains small across the considered settings, whereas variance decreases more clearly with increasing minority-class sample size. These results highlight the finite-sample behavior of AUC estimation under different distributional and sampling conditions. The simulation therefore evaluates the finite-sample behavior of AUC estimation under different distributional and sampling conditions rather than the algebraic identity in (1), which

is exact by definition.

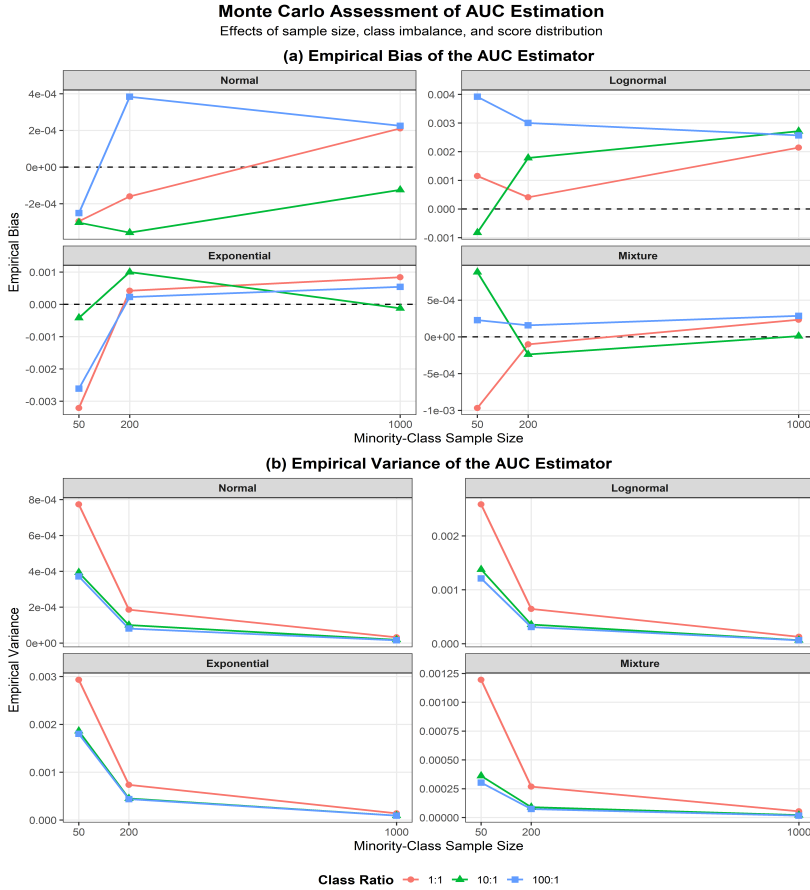


Figure 2: Empirical bias and variance of the AUC estimator across the reported Monte Carlo scenarios.

5.4 Interpretation of results: Income inequality and model performance

To illustrate the distinction between income inequality and ranking-based discrimination, a separate simulated socioeconomic example was considered. A synthetic dataset of 1,000 individuals was used to compare the income-inequality Gini coefficient, AUC of a poverty prediction model, and Cohen's kappa for binary classification agreement.

The estimated income Gini coefficient was 0.379, indicating inequality in the simulated income distribution. This measure reflects distributional concentration and should not be interpreted as the Gini discrimination index associated with ROC analysis.

The prediction model achieved an AUC of 0.744, indicating moderate ranking discrimination, whereas Cohen's kappa at a selected threshold was 0.049. The difference

between these values highlights the distinction between threshold-independent ranking performance and threshold-dependent classification agreement.

Figures 3 and 4 illustrate Lorenz-, Leimkühler-like, and ROC-based representations, demonstrating that these measures provide complementary perspectives rather than measuring the same underlying quantity.

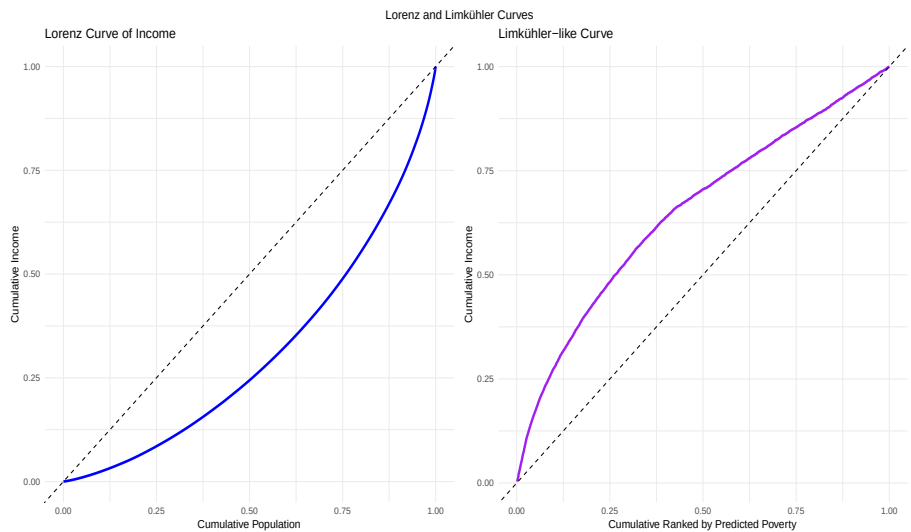


Figure 3: Lorenz curve and Leimkühler-like curve based on model predictions.

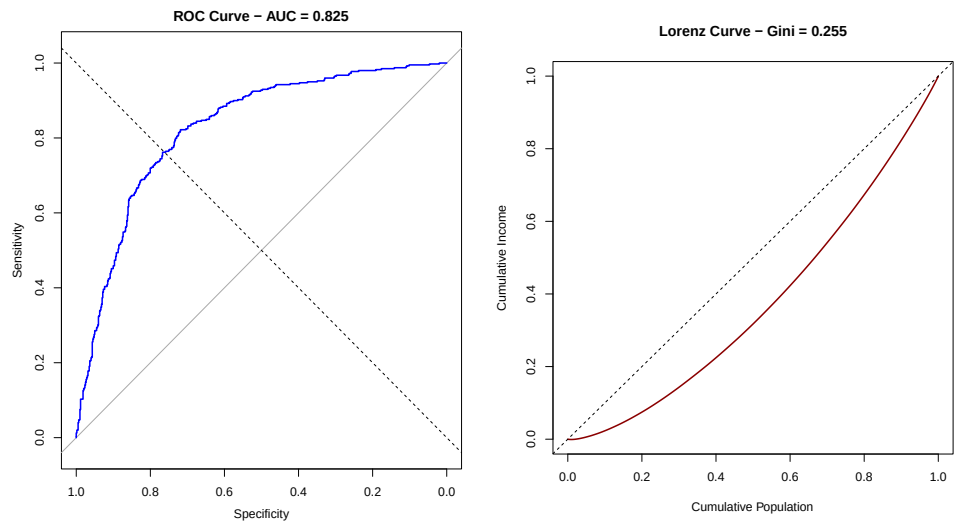


Figure 4: Lorenz Curve and ROC curve for income-based poverty prediction.

5.5 Example: Poverty prediction from socioeconomic features

Consider a simulated dataset of 1,000 individuals with socioeconomic characteristics including education, employment status, and residential region. Income is generated from these characteristics with an error component, and individuals with income below a threshold of 15,000 are classified as poor. A prediction score based on these features is then used to rank individuals according to poverty risk.

The simulated results yield an income-inequality Gini coefficient of 0.255, an AUC of 0.825, and a corresponding Gini discrimination index of 0.65 obtained from (1). The Lorenz curve remains relatively close to the equality line, whereas the ROC curve indicates strong ranking performance. These results demonstrate that income inequality and predictive discrimination can coexist but represent different concepts. The value 0.65 should not be interpreted as economic inequality or compared directly with the income Gini coefficient of 0.255. Instead, it measures the ranking ability of the prediction model, whereas the income Gini coefficient measures concentration in the income distribution. Therefore, Lorenz-based inequality measures and ROC-based discrimination measures may share mathematical structures while retaining distinct interpretations.

5.6 When AUC and Kappa diverge: Theory and practice

Although AUC and Cohen's kappa are both used for evaluating classification performance, they measure different aspects. AUC is a threshold-independent ranking measure that evaluates the probability of correctly ordering positive observations above negative observations, whereas kappa is a threshold-dependent agreement measure based on the resulting confusion matrix.

Cohen's kappa is defined by (3) and depends on observed agreement, chance agreement, the selected classification threshold, and class prevalence. Therefore, a high AUC does not necessarily imply a high kappa value. This distinction is particularly important under class imbalance. A classifier may achieve strong ranking performance while producing low agreement at a specific threshold because of unfavorable sensitivity-specificity trade-offs or changes in marginal class proportions. Consequently, classifiers with similar AUC values may have different kappa values depending on their thresholds and predicted class distributions.

From a practical perspective, AUC is appropriate for assessing overall discrimination independent of a decision threshold, whereas kappa provides complementary information when categorical decisions are required. Under imbalanced settings, both measures should be interpreted together with threshold-dependent quantities such as sensitivity, specificity, and precision, with threshold selection guided by the practical costs of classification errors.

6 Illustrative applications

The following examples illustrate how inequality-based representations and ranking measures can be examined within a common framework. They are not intended as empirical validation of a general relationship between income inequality and classification performance, but rather as controlled examples for comparing cumulative, concentration, and ranking-based measures.

6.1 Simulated digital platform example

A simulated example inspired by income patterns discussed for digital platforms is considered. The objective is not to estimate actual income distributions of specific platforms, but to construct contrasting income distributions representing different levels of heterogeneity (Digikala, 2020; Hall and Krueger, 2017; Zha et al., 2020). A heterogeneous mixture distribution is used to represent greater income concentration, while a more homogeneous distribution represents lower concentration. The analysis combines Lorenz-based measures of income inequality with ROC-based evaluation of the ability of income rankings to distinguish between predefined income groups (bottom 80% versus top 20%).

Figure 5 presents the corresponding Lorenz curves. The more heterogeneous distribution shows greater deviation from the equality line, reflecting higher income concentration. This example should not be interpreted as evidence that income inequality causes better classification performance. Instead, it demonstrates that concentration measures and ranking-based discrimination may vary together under specific data-generating mechanisms while retaining different interpretations. Figure 6 shows the ROC curves obtained from the income-based ranking. Since the groups are defined from different regions of the income distribution, greater heterogeneity leads to stronger separation in this simulated setting.

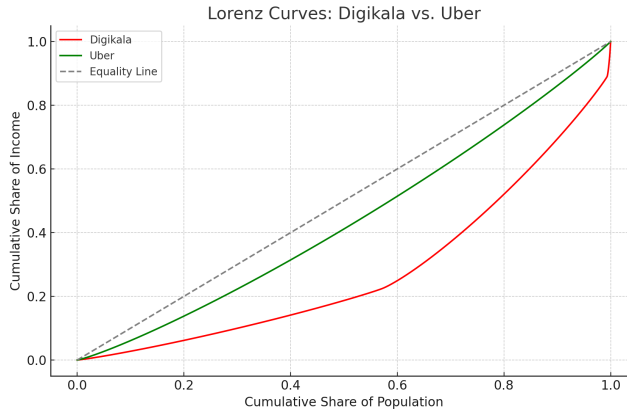


Figure 5: Lorenz curves for simulated income distributions representing contrasting degrees of income concentration.

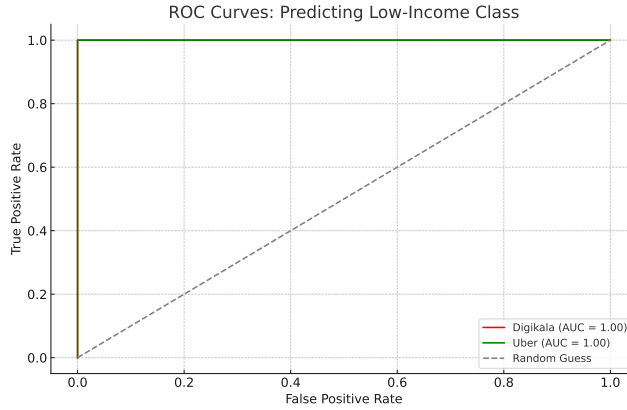


Figure 6: ROC curves for distinguishing predefined income groups in the simulated platform example.

6.2 Breast cancer classification example

To illustrate the distinction between ranking discrimination and probability calibration in a standard classification setting, the Wisconsin Breast Cancer dataset from the `mlbench` package in R statistical software (R Core Team, 2026) was considered. The dataset contains 699 observations; after removing records with missing values, 683 cases remained. Each observation includes nine ordinal cytological features describing cellular characteristics, with the outcome classified as benign or malignant. This example is intended as an illustration of predictive performance evaluation rather than an independent clinical validation study. It demonstrates how ROC-based discrimination measures can be interpreted alongside calibration-related criteria.

6.2.1 Data preparation and experimental dDesign

Records with missing values were excluded, the identifier variable was removed, and the outcome was encoded as a binary variable, with malignant cases represented as one and benign cases as zero. The data were randomly divided into training and test sets using an 80/20 split. Three classification models, including Logistic Regression, Random Forest, and XGBoost, were fitted and evaluated on the held-out test set using ROC-AUC together with calibration-oriented measures to distinguish ranking ability from the quality of predicted probabilities.

6.2.2 Discrimination performance

The ROC curve evaluates the ability of a classifier to rank positive cases above negative cases across different thresholds, while AUC provides a threshold-independent summary of ranking discrimination (Hanley and McNeil, 1982; Fawcett, 2006).

Table 6 summarizes the predictive performance of the three models. All models achieved high discrimination in this illustrative experiment, with AUC values above 0.99. However, high discrimination does not necessarily imply well-calibrated predicted

probabilities. Figure 7 presents the ROC curves of the three models. The results confirm strong ranking discrimination on the test set; however, ROC-AUC alone does not assess probability calibration. Therefore, calibration analysis is required to determine whether predicted probabilities correspond to observed event frequencies.

Table 6: Performance metrics for breast cancer classification models.

Metric	Logistic Regression	Random Forest	XGBoost
Accuracy	0.9632	0.9632	0.9706
AUC	0.9959	0.9944	0.9921
Sensitivity	0.9583	0.9375	0.9792
Specificity	0.9667	0.9778	0.9667
Precision	0.9583	0.9688	0.9592
F1-score	0.9583	0.9529	0.9691
Brier score	0.0284	0.0244	0.0291

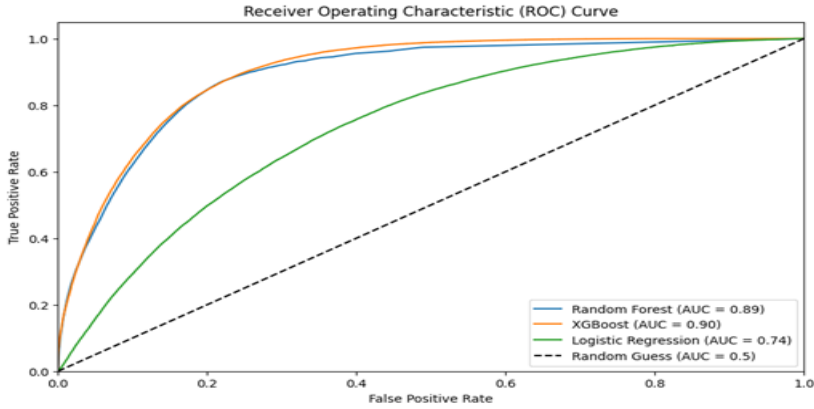


Figure 7: ROC curves for breast cancer classification using Logistic Regression, Random Forest, and XGBoost.

6.2.3 Calibration analysis

Although AUC measures ranking discrimination, it does not assess whether predicted probabilities correspond to observed event frequencies. Therefore, calibration was evaluated using the Brier score:

$$\text{Brier} = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2,$$

where p_i denotes the predicted probability and y_i the observed outcome. Lower values indicate better probability calibration.

The Brier score was further decomposed into reliability, resolution, and uncertainty components:

$$\text{Brier} = \text{Reliability} - \text{Resolution} + \text{Uncertainty}.$$

This decomposition distinguishes calibration from discrimination, where reliability reflects agreement between predicted probabilities and observed frequencies, and resolu-

tion measures separation among groups with different outcome risks. Table 7 summarizes the calibration results. The results show that discrimination and calibration may lead to different model comparisons. A higher AUC does not necessarily imply a lower Brier score or better calibration, highlighting the need to evaluate ranking performance and probability quality separately.

Table 7: Brier score decomposition for the breast cancer classification models.

Model	AUC	Brier Score	Reliability	Resolution
Logistic Regression	0.9959	0.0284	0.0213	0.2190
Random Forest	0.9944	0.0244	0.0063	0.1774
XGBoost	0.9921	0.0291	0.0022	0.1977

6.2.4 Interpretation in relation to the ROC-Gini framework

The breast cancer example illustrates the role of ranking-based measures in a standard classification setting. ROC-AUC evaluates class discrimination, and the corresponding Gini discrimination index follows from the transformation in (1). This quantity should be distinguished from the Lorenz-based income Gini coefficient, which describes inequality rather than predictive ranking performance.

The example further demonstrates that strong discrimination does not necessarily imply well-calibrated probabilities. Therefore, ROC-AUC and calibration measures provide complementary information about predictive models.

Together with the simulated platform example, this application supports the broader framework of the study: Lorenz-based inequality measures and ROC-based discrimination measures may share mathematical structures related to CDFs and ranking, while retaining distinct interpretations and scientific purposes.

6.2.5 Data description and study sample

The breast cancer example was based on the Wisconsin Breast Cancer dataset available in the `mlbench` package in R. The original dataset contains 699 observations, after removing records with missing values, 683 complete cases remained.

Each observation represents a breast cancer case described by nine ordinal cytological features, including clump thickness, uniformity of cell size and shape, marginal adhesion, epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses. The outcome variable indicates benign or malignant status. The identification variable was excluded from model fitting.

The complete cases were randomly divided into training and test sets using an 80/20 stratified split, preserving the class distribution. All reported discrimination and calibration results are based on the held-out test sample. This dataset is used only as an illustrative classification example and should not be interpreted as a clinical validation cohort.

7 Conclusion

This study provided an integrated perspective on the relationships among Lorenz curves, ROC curves, Gini-type measures, concentration representations, and Cohen's kappa. Although these measures share mathematical structures based on CDFs, ordering, and ranking, they quantify different concepts. A key distinction is that the income Gini coefficient measures inequality in a nonnegative distribution, whereas the ROC-based Gini discrimination index measures ranking performance in binary classification. The identity G_{disc} is an established ROC/AUC relationship and should not be interpreted as an equivalence between economic inequality and predictive discrimination. The contribution of this study is therefore methodological: it clarifies the assumptions under which Lorenz-type and ROC-type representations are mathematically connected while preserving their distinct interpretations. Analytical examples and simulations show that these relationships depend on distributional assumptions and finite-sample conditions rather than representing universal equivalences. The illustrative applications further demonstrate that discrimination, agreement, and calibration capture different aspects of predictive performance. ROC-AUC evaluates ranking ability, Cohen's kappa reflects threshold-dependent agreement, and calibration measures assess the consistency between predicted probabilities and observed outcomes. Overall, this study provides a cross-disciplinary framework for interpreting inequality, concentration, ranking, and predictive performance measures without conflating their mathematical relationships with their substantive meanings.

References

- Arnold, B.C. and Sarabia, J.M. (2018). *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*. Vol. 7, New York, NY, USA: Springer.
- Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution*. Cambridge: Cambridge University Press.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, **5**(2):153–163.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1):37–46.
- De Martel, C., Georges, D., Bray, F., Ferlay, J. and Clifford, G.M. (2020). Global burden of cancer attributable to infections in 2018: A worldwide incidence analysis. *The Lancet Global Health*, **8**(2):e180–e190.
- DeSantis, C.E., Ma, J., Gaudet, M.M., Newman, L.A., Miller, K.D., Goding Sauer, A., Jemal, A. and Siegel, R.L. (2019). Breast cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, **69**(6):438–451.
- Dianati-Nasab, M., Salimifard, K., Mohammadi, R., Saadatmand, S., Fararouei, M., Hosseini, K.S., Jiavid-Sharifi, B., Chaussalet, T. and Dehdar, S. (2024). Machine

- learning algorithms to uncover risk factors of breast cancer: Insights from a large case-control study. *Frontiers in Oncology*, **13**:1276232.
- Digikala, (2020). *Performance Report of Sellers*. Retrieved from <https://www.digikala.com/mag/digikala-sellers-income>.
- Erdreich, L.S. (1981). Use of relative operating characteristic analysis in epidemiology: A method for dealing with subjective judgement. *American Journal of Epidemiology*, **114**(5):649–662.
- Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, **542**(7639):115–118.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, **27**(8):861–874.
- Figini, S., Gigliarano, C. and Muliere, P. (2012). Some results on stochastic comparisons of ROC curves. In *46th Scientific Meeting of The Italian Statistical Society*, 1–4.
- Goodenough, D.J., Rossmann, K. and Lusted, L.B. (1974). Radiographic applications of receiver operating characteristic (ROC) analysis. *Radiology*, **110**(1):89–95.
- Hall, J.V. and Krueger, A.B. (2017). An analysis of the labor market for Uber’s driver-partners in the United States. *ILR Review*, **71**(3):705–732.
- Hand, D.J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, **77**(1):103–123.
- Hanley, J.A. and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**(1):29–36.
- Hanley, J.A. and McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, **148**(3):839–843.
- Irwin, R.J. and Hautus, M.J. (2015). Lognormal Lorenz and normal receiver operating characteristic curves as mirror images. *Royal Society Open Science*, **2**(2):140280.
- Gasparini, M. and Sacchetto, L. (2020). On the definition of a concentration function relevant to the ROC curve. *Metron*, **78**(3):271–277.
- Gasparini, M. and Sacchetto, L. (2023). Concentration and ROC curves, revisited. *Sankhya A*, **85**(1):292–305.
- Gini, C. (1914). Measurement of inequality of incomes. *The Economic Journal*, **31**(121):124–125.
- Hardt, M., Price, E. and Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, **29**:3315–3323.
- Kämpke, T. and Radermacher, F.J. (2015). *Income Modeling and Balancing*. Lecture Notes in Economics and Mathematical Systems.

- Kaymak, U., Ben-David, A. and Potharst, R. (2012). The AUK: A simple alternative to the AUC. *Engineering Applications of Artificial Intelligence*, **25**(5):1082–1089.
- Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons.
- Liu, C., White, M. and Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography*, **34**(2):232–243.
- Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**(70):209–219.
- Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, **8**(4):283–298.
- Pontius, R.G., Jr. and Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, **32**(15):4407–4429.
- R Core Team (2026). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org>.
- Song, X., Chu, J., Guo, Z., et al. (2024). Prognostic prediction of breast cancer patients using machine learning models: A retrospective analysis. *Gland Surgery*, **13**(9):1575–1587.
- Schechtman, E. and Schechtman, G. (2019). The relationship between Gini terminology and the ROC curve. *Metron*, **77**(3):171–178.
- Zha, L., Yin, Y. and Du, Y. (2020). Surge pricing and labor supply in the ride-sourcing market. *Transportation Research Part B: Methodological*, **139**:381–400.