POLITECNICO
MILANO 1863

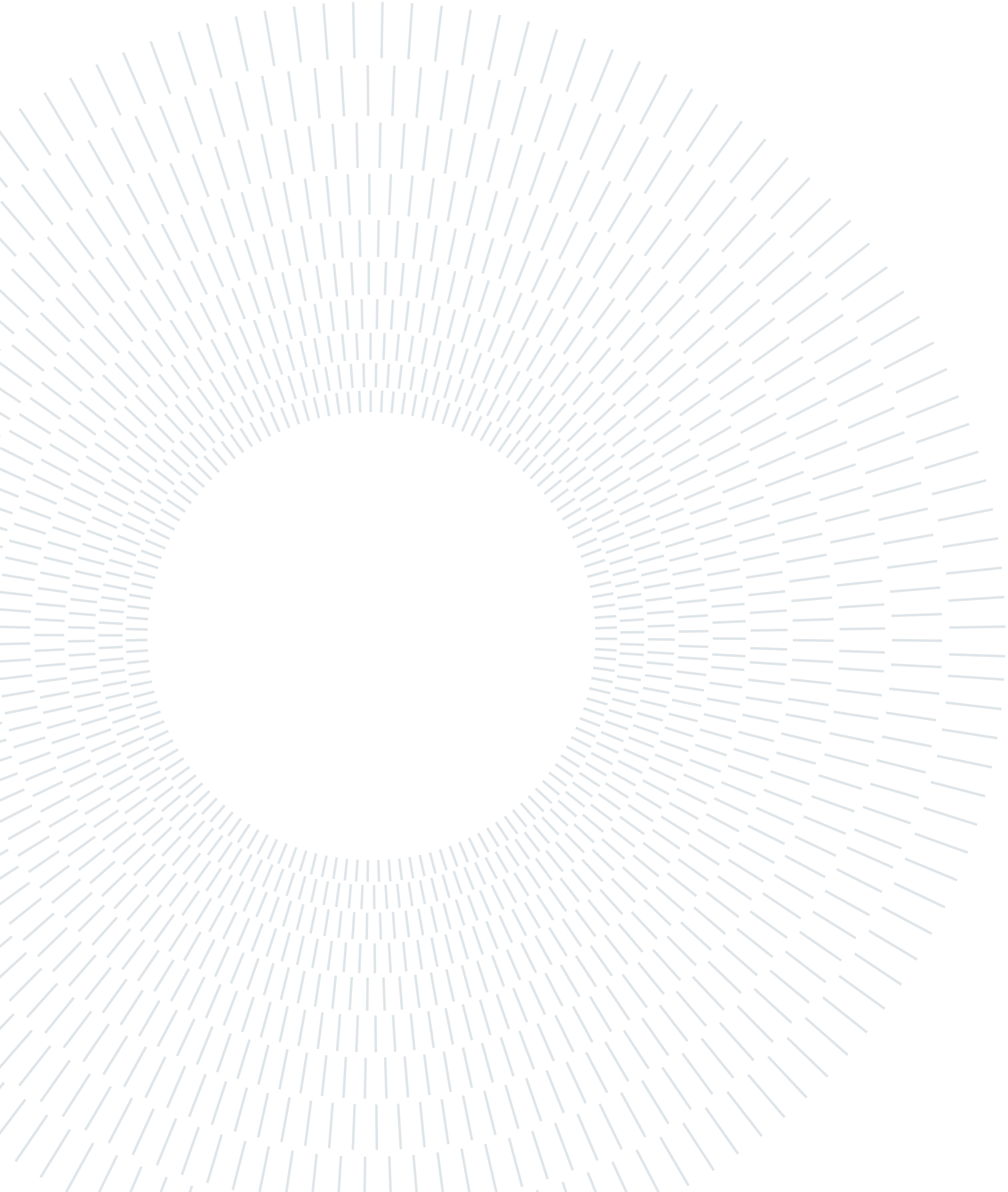
SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

A Theory-Driven Approach to Large Language Models Alignment with Human Feedback

Tesi di Laurea Magistrale in
Computer Science and Engineering -
Ingegneria Informatica

Author: **Michele Simeone**

Student ID: 221403
Advisor: Prof. Alberto Maria Metelli
Co-advisors: Tommaso Bianchi, Simone Drago, Gianmarco Genalti
Academic Year: 2023-24



Abstract

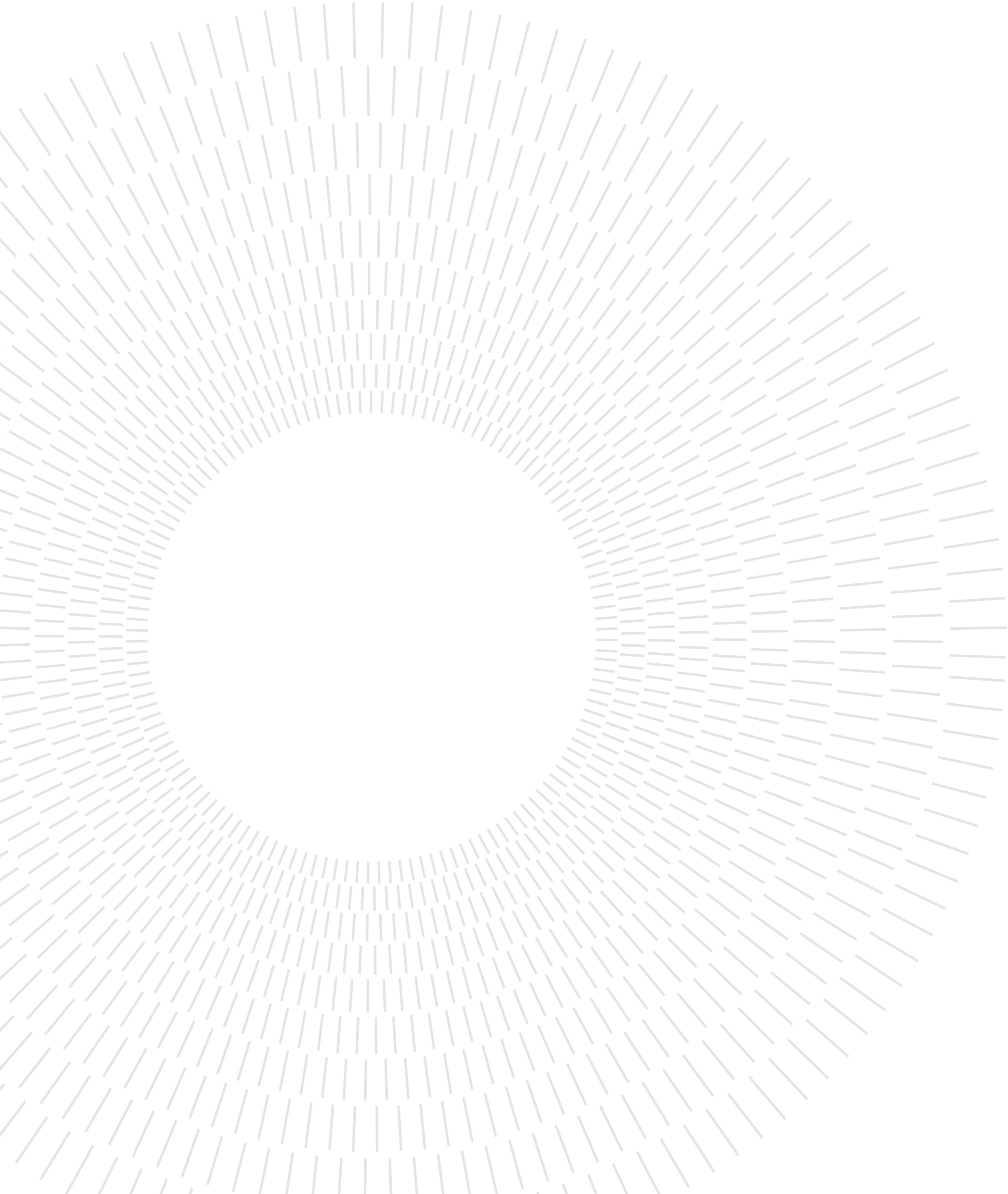
This thesis investigates the response alignment problem for Large Language Models (LLM). Given two possible answers to a user query, the alignment problem consists of suggesting the alternative closest to the end user's preferences, avoiding suggesting incorrect or harmful answers.

The goal of this thesis is to produce an algorithm capable of carrying out this process by emphasizing a cost-efficient solution from a computational point of view and in the number of samples employed.

To achieve the objective, we move a step of abstraction and formulate the alignment problem as an offline linear feasibility problem. In this problem, each answer is associated with a vector of contexts which contains the values given to a set of properties that fully characterize a response, such as length or clarity. The contexts vectors combined with the user's preference constitute the constraints of the problem. The solution is an area of feasibility that summarizes the user's inclination towards the context. Consequently, once the problem is solved, we can directly suggest to the user answers that reflect his tastes, generating them according to the identified area. We then conduct a theoretical analysis in three distinct scenarios, where the unknown distributions governing the LLM's context vectors follow an isotropic Gaussian, a non-isotropic Gaussian, or a binomial distribution. We establish a theoretical guarantee of the sample complexity for each case. Lastly, we validate our algorithm using a real dataset.

Experimental results show that compared to a baseline majority voting approach, in which the suggestion is chosen as the mode of the suggestions of 31 LLMs, our solution is more promising and correctly identifies the user's preference.

Keywords: Alignment, LLM, Offline Feasibility Problem, Sample Complexity



Abstract in Lingua Italiana

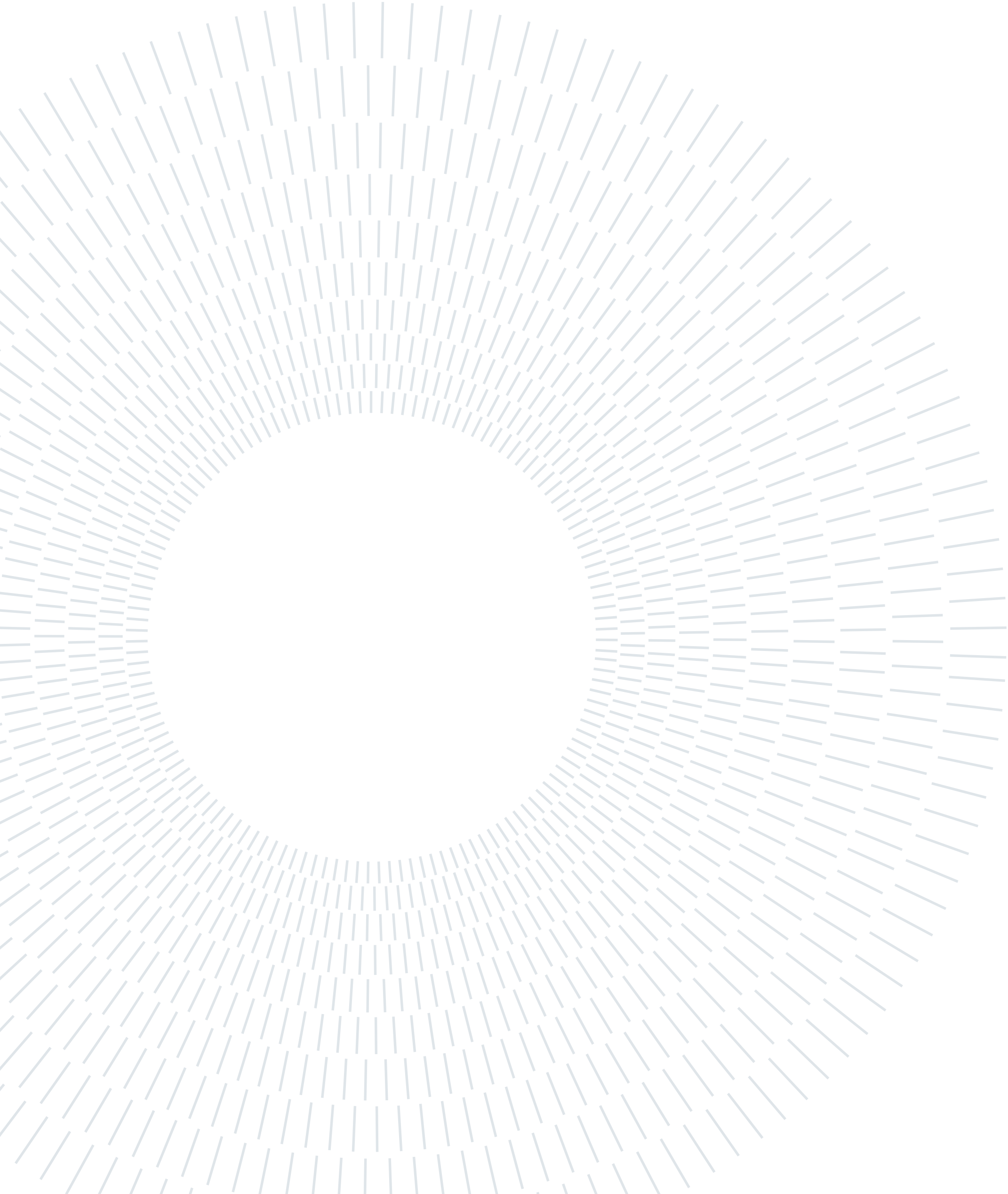
Questa tesi indaga il problema dell'allineamento delle risposte per i Large Language Models (LLM). Date due possibili risposte ad una domanda dell'utente, il problema dell'allineamento consiste nel suggerire all'utente finale l'alternativa che più si avvicina alle sue preferenze, evitando di suggerire risposte errate o dannose.

L'obiettivo di questa tesi è quello di produrre un algoritmo in grado di effettuare questo processo enfatizzando una soluzione economica dal punto di vista computazionale e in termini di numero di campioni utilizzati.

Per raggiungere l'obiettivo, facciamo un passo di astrazione e formuliamo il problema dell'allineamento come un problema di feasibility lineare offline. In questo problema ad ogni risposta è associato un vettore di contesti che contiene i valori dati ad un insieme di proprietà che caratterizzano pienamente ogni risposta, quali la lunghezza o la chiarezza. I vettori dei contesti combinati con le preferenze dell'utente costituiscono i vincoli del problema. La soluzione è un'area di feasibility che sintetizza la propensione dell'utente verso gli specifici contesti. Di conseguenza, una volta risolto il problema, possiamo direttamente suggerire le risposte che rispecchiano i gusti dell'utente, generandole in accordo con l'area identificata. Conduciamo quindi un'analisi teorica in tre scenari distinti, in cui le distribuzioni sconosciute che governano i vettori di contesto del LLM seguono una distribuzione gaussiana isotropa, gaussiana non isotropa o una binomiale. Per ciascun caso stabiliamo una garanzia teorica sulla complessità del campione. Infine, valutiamo l'algoritmo proposto utilizzando un set di dati reale.

I risultati sperimentali mostrano che rispetto a un approccio baseline di voto a maggioranza, in cui la risposta suggerita è la moda dei suggerimenti di 31 LLM, la nostra soluzione è più promettente e identifica correttamente la preferenza dell'utente in diversi scenari.

Parole chiave: Allineamento, LLM, Problema di Feasibility Offline, Complessità del Campione



Contents

1 | Introduction

Decision-making in Artificial Intelligence (AI) often requires selection among multiple potential outcomes, each of which carries distinct implications and trade-offs [1]. The capability to make optimal selections among competing alternatives is of paramount importance across numerous domains, including autonomous systems [2], recommendation engines [3], medical diagnostics [4], financial modeling [5], and Natural Language Processing (NLP) [6]. A critical point in this process is the integration of alignment mechanisms, ensuring that AI-driven decisions adhere to human values and expectations.

One of the most established and widely recognized applications of multi-output selection in AI is within recommendation systems, as exemplified by e-commerce platforms and streaming services [3]. When a user searches, e.g., for a product or a movie, the system must not only rank but also present the most pertinent options. The challenge is balancing competing goals such as relevance, diversity, and personalization to improve user engagement. Human feedbacks acquired through user interactions serve to iteratively refine these recommendations over time, therefore become fundamental.

In the domain of medical diagnostics, AI-assisted systems frequently encounter the necessity to choose between multiple diagnostic hypotheses. Given a set of symptoms and corresponding test results, an AI model may propose multiple plausible conditions that explain the observed data [4]. The selection between these hypotheses becomes of vital importance, as it can have significant consequences on patients, leading to the choice of different treatment strategies.

Another well-known application in the use of AI decision-making systems can be found in the financial sector [5], where in algorithmic trading, AI systems continuously evaluate multiple trading strategies in response to dynamic market conditions and associated risk assessments. There is a trade-off in this industry between profitability and risk mitigation, demanding sophisticated optimization techniques and robust predictive modeling.

In the NLP field, Chatbots and LLMs often generate multiple potential responses to a given user query, and the most appropriate one must be selected based on contextual understanding and predefined objectives. For instance, in addressing a complex inquiry, an

AI system may generate one response that prioritizes accuracy and another that emphasizes simplicity. New paradigms, such as Reinforcement Learning from Human Feedback (RLHF) [6], play a key role in ensuring alignment with user values.

In general, the selection of alternative outputs constitutes a fundamental challenge in AI applications that necessitate a judicious balance between algorithmic optimization and alignment with human-centered values.

Alignment of Large Language Models

LLLMs represent a specialized category of AI systems designed to process and generate human-like text through the utilization of extensive training data. These models have found widespread application within NLP tasks, including but not limited to text classification [7], translation [8], summarization[9], and conversational AI [10]. As computational capabilities advance and training methodologies become increasingly sophisticated, LLMs continue to proliferate across a diverse array of domains, spanning customer service [11], healthcare [12], legal analysis [13], software development [14], education [15], and content generation [16].

Early iterations of LLMs predominantly relied on rule-based systems and statistical models, which, while ensuring relative ease of alignment, constrained their flexibility and adaptability. However, with the advent of deep learning and transformer-based architectures [17], contemporary models show substantially improved capabilities, allowing them to understand contextual nuances, reason over extensive information, and dynamically adapt to different linguistic styles.

Currently, research in the field of LLMs predominantly focuses on two related objectives improving model efficiency and refining alignment strategies. The first goal involves the development of techniques such as model pruning, knowledge distillation [18], and adaptive inference that mitigate computational costs and reduce environmental impact. On the other hand, alignment strategies seek to ensure that models produce outputs that remain real, unbiased, and ethically valid [19].

The alignment of LLMs is an ongoing challenge necessitating a multidisciplinary approach that integrates technical advancements, ethical considerations, and sustained human oversight. Notable alignment methodologies include RLHF, adversarial robustness testing and prompt engineering.

Motivation

With the advancement of increasingly powerful LLMs, the financial and computational resources required to align high-performance models have become prohibitive for all but a few organizations, which rely on executing techniques that have no theoretical guarantees that they will work. This implies the need to explore alternative approaches to align LLM, which leverages response characteristics as a source to perform this process and provide associated sample complexity.

Goal

The primary objective of this thesis is to understand the theoretical aspects of the LLM alignment problem, providing an algorithm with theoretical guarantees and good experimental performance at least in a simplified scenario. We focus on the study of the sample complexity associated with the proposed approach and empirically evaluating its effectiveness on real-world data.

Original Contribution

The principal contribution of this research is the development of a novel framework for aligning the selection process between two options with the preferences of a target entity, specifically in the context of aligning LLMs. While numerous existing approaches have been proposed for LLM alignment this work introduces an alternative that has theoretical guarantees. To do this we formulate the alignment problem as an offline feasibility problem, eliminating the need for iterative online interactions typical of the solutions proposed in the state of art. This setting allows us to formalize a sample complexity for our algorithm.

1.1. Thesis Outline

This thesis is structured into the following five chapters, each addressing a key aspect of the research.

In Chapter 2, we provide the technical tools that we will use during the thesis and then present a review of the state of the art of LLM alignment, delivering a complete overview of existing approaches and methodologies.

In Chapter 3, we introduce the mathematical modeling and formal problem formulation. We define the problem setting in terms of inputs, outputs, and associated constraints.

Specifically, we model the offline feasibility problem, where the constraints are represented by the contexts associated with the two inputs, and the output is an estimator of the target's preference.

In Chapter 4, we present our proposed algorithm for aligning a LLM and derive the sample complexity associated with it across three different settings. Initially, we define the theoretical framework within which our algorithm operates. Subsequently, we provide a formal characterization of sample complexity for each scenario.

Chapter 5 details the experimental evaluation of the proposed solutions. We first describe the experimental setup, then we discuss the results in depth, analyzing both the quantitative and qualitative aspects of them.

Finally, in Chapter 6, we summarize the contributions of this thesis, outlining the strengths and limitations of the proposed methodology. We also discuss potential directions for future research, suggesting possible extensions and improvements to our approach.

2 | Related Works

In this chapter we explore existing research and methodologies that are closely related to the problems and solutions addressed in this thesis.

We begin by examining the problem of selecting the optimal alternative, a decision challenge that has been studied in fields such as decision theory, machine learning, and economics. Various models are discussed, each offering different trade-offs in terms of efficiency, interpretability, and applicability in uncertain environments.

Next, we analyze the offline linear feasibility problem, a computational problem where the goal is to determine the existence of feasible solutions within a system of linear constraints. We present widely used solution techniques highlighting their respective advantages and limitations in efficiently solving the problem.

Additionally, we provide a comprehensive review of state-of-the-art methodologies for aligning LLMs with human feedback. This section delves into recent advances in RLHF and Supervised Learning approaches such as Directed Preference Optimization (DPO). We describe the main challenges in model alignment, including trade-offs between utility and harmlessness, computational costs, and inherent difficulties associated with modeling human preferences.

By presenting an overview of these related works, this chapter establishes the necessary theoretical and methodological background that informs the subsequent contributions of this thesis.

2.1. Related Problems and Solutions

2.1.1. Selection of the Optimal Alternative

The problem of selecting the optimal alternative between two given options, each characterized by an associated feature vector, has been extensively examined across multiple disciplines, including decision theory, machine learning, and economics [20]. Early seminal contributions in this field primarily investigated the computational complexity of decision-making processes and formulated foundational selection criteria.

In recent years, the development of advanced algorithms capable of identifying the optimal (or near-optimal) alternative based on multidimensional feature representations has gained significant attention. Contemporary literature typically frames this problem as a multi-criteria decision-making (MCDM) challenge, wherein different attributes must be systematically evaluated to determine the most favorable choice [21].

The selection process between two alternatives can be approached using various methodological paradigms, including deterministic and probabilistic models. This section provides a concise overview of key decision-making methodologies proposed in recent research, spanning from traditional heuristic-based frameworks to probabilistic models.

Heuristic-Based Decision-Making

Heuristic-based decision-making methods provide an efficient mechanism for selecting between two alternatives. These approaches exploit approximations and intuitive reasoning rather than exhaustive optimization techniques, and are therefore suitable for environments full of uncertainty.[22].

A common heuristic approach is the Weighted Sum Model (WSM), which ranks alternatives by computing a weighted sum of feature values. Given an alternative A_i, its score S_i is computed as:

$$S_i = \sum_{j=1}^{n} w_j x_{ij} \ , \tag{2.1}$$

where w_j represents the weight assigned to each feature and x_{ij} denotes the corresponding feature value. The alternative with the highest computed score is selected [23]. Although computationally efficient and easy to interpret, this approach assumes linearity and is highly sensitive to weight assignments.

Another structured heuristic method is the Analytic Hierarchy Process (AHP), which decomposes decision-making into a hierarchical framework [24]. Pairwise comparisons are conducted to determine the relative importance of features, leading to the construction of a comparison matrix $\mathbf{A}$, from which priority weights $\mathbf{w}$ are derived through the computation of the eigenvector corresponding to the largest eigenvalue λ_{max}:

$$\mathbf{A}\mathbf{w} = \lambda_{\text{max}}\mathbf{w} \ . \tag{2.2}$$

This method provides a structured evaluation framework and mitigates bias but is computationally more demanding than simpler heuristic approaches. While WSM offers a rapid and interpretable decision-making process, AHP proves advantageous in structured assessments requiring expert judgment. The selection of an appropriate heuristic approach depends on factors such as problem complexity, domain expertise, and computational constraints.

Probabilistic Decision Models

Probabilistic decision models facilitate the selection between two alternatives by estimating the probability of each option being optimal and subsequently choosing the alternative with the highest likelihood.

A well-known probabilistic approach is logistic regression, which estimates selection probabilities using a sigmoid function. Given a feature vector $\mathbf{X} = (x_1, ..., x_d) \in \mathbb{R}^d$, the probability that an alternative is chosen is modeled as:

$$\mathbb{P}(y = 1|\mathbf{X}) = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{X}+b)}} \ , \tag{2.3}$$

where, $\mathbf{w} = (w_1, w_2, ..., w_d) \in \mathbb{R}^d$ represents the weight vector, where each coefficient w_i captures the contribution of the corresponding feature x_i to the decision boundary. The bias term $b \in \mathbb{R}$ serves as an offset, allowing flexibility in defining the decision boundary independently of the feature values. The expression $\mathbf{w}^T\mathbf{X} + b$ defines the log-odds of selecting alternative $y = 1$, which is then transformed into a probability by the sigmoid function.

This method is interpretable, computationally efficient, and provides probability estimates [25]. However, it relies on the assumption of a linear relationship between features and targets.

Another probabilistic approach is the Naïve Bayes classifier, which leverages Bayes' theorem under the assumption of feature independence. The probability of an outcome y given a feature set $\mathbf{X} = (x_1, ..., x_d) \in \mathbb{R}^d$ is computed as:

$$\mathbb{P}(y|\mathbf{X}) \propto \mathbb{P}(y) \prod_{i=1}^{d} \mathbb{P}(x_i|y) \ . \tag{2.4}$$

This method is computationally efficient and performs well on small datasets while being robust to irrelevant features. However, its assumption of feature independence can be restrictive in practical applications.

Probabilistic decision models play a crucial role in decision-making under uncertainty.

While logistic regression is preferred for its interpretability and efficiency in linearly separable problems, Naïve Bayes proves advantageous for small datasets with independent features. The selection of an appropriate probabilistic model depends on data complexity, dataset size, and the degree of feature interdependence [26].

2.1.2. Offline Linear Feasibility Problem

The offline linear feasibility problem constitutes a fundamental class of computational problems wherein the objective is to assess the existence of at least one point that satisfies a given system of linear constraints [27]. Formally, given a constraint system of the form:

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} \ , \tag{2.5}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ denotes a coefficient matrix, $\mathbf{x} \in \mathbb{R}^n$ represents the vector of decision variables, and $\mathbf{b} \in \mathbb{R}^m$ is the constraint vector, the task is to determine whether there exists an assignment of $\mathbf{x}$ that satisfies all constraints.

Linear Programming as a Feasibility Approach

A widely adopted approach to addressing linear feasibility problems is through Linear Programming (LP), which, instead of merely verifying feasibility, often seeks to optimize a specified objective function [28]. In its standard formulation, a linear programming problem is expressed as follows:

$$\max \mathbf{c}^T\mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \ , \tag{2.6}$$

where $\mathbf{c} \in \mathbb{R}^n$.

If this optimization problem admits a finite optimal solution, then it follows that the corresponding feasibility problem possesses a feasible solution. Conversely, infeasibility in the optimization framework implies the nonexistence of any satisfying solution for the feasibility problem.

Among the methodologies for solving linear feasibility problems, the two most used paradigms are the Simplex method [29] and the Interior Point method [30]. These approaches exhibit distinct properties that make them suitable for different problem cases.

The Simplex Method

The simplex algorithm represents a classical and widely utilized technique for solving linear programming problems, including feasibility determination [29]. It systematically navigates the vertices of the feasible polyhedron to either identify an optimal solution or establish infeasibility. It is particularly advantageous for small problems due to its ability to provide exact solutions. However, despite its empirical efficiency, the simplex method exhibits an exponential worst-case time complexity that is $\mathcal{O}(2^n)$. Under certain degeneracy conditions, it may also encounter cycling [31].

Interior-Point Method

An alternative to the simplex approach is provided by interior point methods, which has gained importance due to its reduced polynomial complexity of $\mathcal{O}(n^{3.5}L)$ (where L is the number of bits of the solution) and their effectiveness in dealing with large-scale optimization problems [30]. However, its implementation is more complex.

In conclusion, the offline linear feasibility problem constitutes a fundamental challenge in optimization and computational mathematics. While the simplex method offers interpretability and exact vertex-based solutions, interior-point methods provide robust performance for large-scale instances with enhanced numerical stability. The selection of an appropriate algorithm is therefore contingent upon the specific characteristics of the feasibility problem at hand [32].

2.2. State of Art Review

2.2.1. Aligning an LLM

The task of aligning Machine Learning (ML) models with human feedback has been a major focus of research since the early 2000s [6]. Pioneering research in this area has focused primarily on elucidating the difficulties associated with learning from human preferences.

Lately, considerable progress has been made in developing algorithms that can align LLM with human feedback. These advances take advantage of Reinforcement Learning (RL), supervised fine-tuning, and scalable oversight techniques [33]. Contemporary research conceptualizes the alignment problem as an optimization challenge that seeks to balance model performance with interpretability and ethical considerations.

The alignment of models with human feedback can also be approached through explicit reward modeling or implicit preference learning. Additionally, alignment methodologies can be categorized into online and offline approaches, enabling models to be continuously updated based on newly acquired human inputs [34].

This section provides a concise overview of the alignment methodologies proposed in recent years. The surveyed approaches range from traditional RLHF to supervised learning techniques, aiming to enhance alignment while mitigating biases and inconsistencies.

Reinforcement Learning from Human Feedback (RLHF)

RLHF [6] constitutes an advanced methodology within the domain of ML that integrates human evaluative input to enhance the processes of computational models.

In detail, RL is based on the development of decision-making algorithms that seek to maximize an objective function, thus guaranteeing the derivation of highly precise and optimized outcomes. RL operates within a setting in which an agent iteratively interacts with an environment, observing a state, and taking an action that yields a reward, thereby learning an optimal policy through exploration and exploitation strategies [35]. Incorporation of human feedback into the reward function of RL systems has demonstrated considerable efficacy, particularly in the domain of AI model alignment, where ensuring the adherence to human values and expectations remains a significant challenge [6]. By leveraging human preferences, RLHF circumvents the limitations of purely automated reward functions, which may struggle to capture nuanced ethical or contextual considerations.

The process involves presenting annotators with example queries paired with two distinct responses, from which they select the response that best aligns with given instructions. These human-labeled preferences form the basis for generating a dataset that encodes nuanced distinctions that purely algorithmic methods might overlook. The resulting dataset is then used to train a reward model, which serves as the reward function for fine-tuning the LLM using different algorithms, among which the most used currently is the Proximal Policy Optimization (PPO). PPO is favored due to its stability and efficiency in policy updates, allowing for iterative refinements that gradually improve alignment with human-preferred outputs [36]. This iterative feedback loop between human annotators and RL-based fine-tuning constitutes a powerful paradigm for developing AI models that are both robust and aligned with human expectations.

Models such as InstructGPT, which follow this methodology, have demonstrated superior performance compared to the original GPT-3 model, despite utilizing significantly fewer parameters [33]. However, achieving both increased correctness and reduced toxicity remains a challenging task, as the quality of the model's outputs is highly dependent on the instructions and objectives provided to annotators during training. This is primarily due to the inherent tension between helpfulness and harmlessness. To address this issue, some models incorporate dual annotation teams, one tasked with selecting the most useful and honest responses, while the other evaluates the most harmful responses [37].

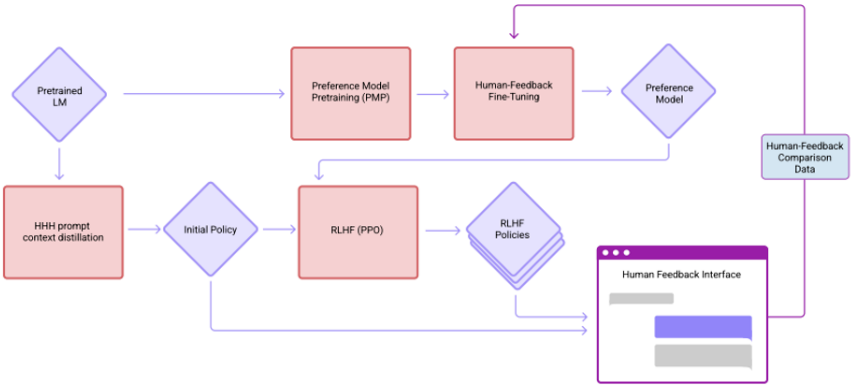

Pretrained LM
Preference Model Pretraining (PMP)
Human-Feedback Fine-Tuning
Preference Model
Human-Feedback Comparison Data
HHH prompt context distillation
Initial Policy
RLHF (PPO)
RLHF Policies
Human Feedback Interface

Figure 2.1: Illustration of the fine-tuning loop of HHGPT [37].

Additionally, RLHF models can operate in offline settings by replacing PPO, which has been observed to be unstable for language fine-tuning, with a filtering-based alignment system. This approach incorporates filtering alignment to pre-screen samples, reward-weighted regression to modify loss functions and emphasize impactful samples, and conditional alignment to enhance adherence to human preferences. These methods achieve strong results despite using about 9% of the computational resources needed for PPO approaches. However, a drawback is the out-of-distribution issue, which arises because models are trained on offline datasets that lack the full diversity of real-world data [38].

Supervised Learning

Implicit preference learning involves refining the models to better capture human preferences without relying exclusively on explicit feedback. This approach enhances model performance by subtly integrating human-like judgments into the training process.

A notable method in this domain is Direct Preference Optimization (DPO), which simplifies the alignment process by directly adjusting the model based on human preferences, thereby eliminating the need for a separate reward model or complex reinforcement learning loops. DPO formulates alignment as a supervised learning problem over preference data, while achieving results comparable to traditional methods [39].

Given a dataset of response pairs (x, y^+, y^-), where x represents the input or prompt provided to the model, y^+ denotes the preferred response, and y^- corresponds to the less preferred response, DPO optimizes the objective function:

$$L(\theta) = \log \frac{\pi_\theta(y^+|x)}{\pi_\theta(y^-|x)} \ , \tag{2.7}$$

where π_θ is the policy model, assigning probabilities to responses given an input x, and θ represents its parameters, which are optimized to favor preferred responses. This objective function encourages the model to assign higher probabilities to preferred responses without necessitating an explicit reward model.

Although these algorithms are extensively employed in the fine-tuning process of a LLM, they exhibit several critical limitations. In particular, the procedure for obtaining feedback constitutes a significant challenge, as it is not only computationally expensive but also prone to issues related to data quality. Moreover, the alignment of the annotators' judgments with the intended objectives remains suboptimal and difficult to accurately assess, thereby introducing potential inconsistencies in the fine-tuning process [40].

3 | Problem Formulation and Methods

In this chapter, we introduce the problem formulation and our methodological framework for preference learning. We begin by defining the problem of determining human preference between two given responses, each represented as a feature vector in an n-dimensional space. A mathematical function is used to quantify the alignment of each response with a human decision-maker's preference vector, and a comparison framework is established to determine which response is preferred.

We proceed by discussing the offline feasibility problem, which aims to determine a preference vector that best explains a given dataset of pairwise comparisons. We frame the problem as a set of constraints on the preference vector and then, we analyze the statistical properties of response feature vectors and their differences.

To facilitate the analysis, several key assumptions are introduced. These assumptions enable a structured approach to modeling human preferences and deriving theoretical guarantees.

3.1. Interaction Protocol

Our interaction protocol begins with a user providing a query Q, in response to which two possible answers are generated: A_1 and A_2. In general, the entities indicated as Q, A_1 and A_2 can correspond to objects of different nature such as text, images and even music files. However, within the scope of this study, these elements will be conceptualized as an interaction with an LLM language used as a chatbot. Each answer is also associated with a vector, $\mathbf{c}_1$ for A_1 and $\mathbf{c}_2$ for A_2 both belonging to $\mathbb{R}^n$. They represent the properties of the corresponding responses and are called *contexts*, these can be easily extracted from the responses using existing models. Specifically, considering n different properties of the answers (for example the length or the stylistic register), the j-th element of the vector represents a score for that category called $[c_{i,1}, \ldots, c_{i,n}]$ with $i \in \{1, 2\}$. The set of contexts fully characterizes each response. Note that the user who provides the preference cannot explicitly observe either $\mathbf{c}_1$ or $\mathbf{c}_2$.

Furthermore, we consider a unknown preference vector $\mathbf{v}^* \in \mathbb{R}^n$, which represents the preference for each context of the human decision-maker. Each element of this vector, v_i^*, can assume a value within the interval $[-1, 1]$.

We assume that there exists a function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, which, given an input vector $\mathbf{c}$ (representing a response) and the preference vector $\mathbf{v}^*$, returns a numerical score. This score represents how well the response characterized by $\mathbf{c}$ aligns with human preferences as described by $\mathbf{v}^*$. The comparison of the scores $f(\mathbf{c}_1, \mathbf{v}^*)$ and $f(\mathbf{c}_2, \mathbf{v}^*)$ determines which response the human would prefer, with three possible outcomes:

- $f(\mathbf{c}_1, \mathbf{v}^*) < f(\mathbf{c}_2, \mathbf{v}^*)$ indicating a preference for response A_2.
- $f(\mathbf{c}_1, \mathbf{v}^*) > f(\mathbf{c}_2, \mathbf{v}^*)$ indicating a preference for response A_1.
- $f(\mathbf{c}_1, \mathbf{v}^*) = f(\mathbf{c}_2, \mathbf{v}^*)$ indicating indifference between the two responses.

Our objective is to identify a feasible region $\mathbf{V} \in \mathbb{R}^n$ within the n-dimensional hyperspace that contains the preference vector $\mathbf{v}^*$.

Algorithm 3.1 Iterative Interaction Protocol

```
Ensure: Identification of feasible preference region $\mathbf{V}$
Initialization: Define the feasible region $\mathbf{V} \subset \mathbb{R}^n$
while feasible region $\mathbf{V}$ is not sufficiently small do
    STEP 1: Generate responses
    USER provides query $Q$
    LLM generates two responses $A_1$ and $A_2$
    LLM extracts context vectors $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^n$
    STEP 2: Collect preference
    USER provides preference $p$
    STEP 3: Update feasible region
    $\mathbf{V} \leftarrow$ update_feasibility_region($\mathbf{V}, \mathbf{c}_1, \mathbf{c}_2, p$)
end while

return Identified preference region $\mathbf{V}$
```

3.2. Offline Feasibility Problem

To define the feasibility problem rigorously, we introduce the necessary notation and underlying structure. Let $\mathbb{D} = \{(\mathbf{c}_1^i, \mathbf{c}_2^i, p_i)\}_{i=1}^k$ be a dataset consisting of k samples, where each sample i is associated with two feature vectors, $\mathbf{c}_1^i$ and $\mathbf{c}_2^i$. Additionally, each sample is accompanied by a preference label p_i, which takes a binary value in $\{0, 1\}$, indicating the preferred option between the two, specifically if the $p_i = 0$ the first option is preferred, otherwise the second is the chosen one. There are also two unknown distributions $\mathcal{G}, \mathcal{Z}$ such that:

$$\mathbf{c}_1 \sim \mathcal{Z}, \mathbf{c}_2 \sim \mathcal{G} \ ,$$

however, the preferences associated with each row are deterministic.

The goal is to determine a preference vector $\hat{\mathbf{v}} \in \mathbb{R}^n$ that best fits a given set of pairwise comparisons. The problem is structured so that the preference vector $\hat{\mathbf{v}}$ must satisfy constraints that encode the preference relationships observed in the data. Specifically, suppose a given preference label is $p_i = 0$, the difference of the application of f with context vectors and $\mathbf{v}^*$ must be nonnegative, ensuring that the first vector is at least as preferred as the second. Mathematically, this is expressed as:

$$p_i = \mathbb{1}_{\{f(\mathbf{c}_1, \mathbf{v}^*) - f(\mathbf{c}_2, \mathbf{v}^*) < 0\}} \ .$$

The feasibility problem is formulated by enforcing these conditions as finding $\hat{\mathbf{v}}$ that best satisfies the observed preference relationships.

3.3. Model Assumptions

The learning protocol introduced in the previous section is very general. In this section, we introduce a set of assumptions to delimitate the focus of this work while preserving its applicability to real-world scenarios. Each assumption reflects the problem's underlying structure and defines the preference function's mathematical properties, ensuring a reusable theoretical background to provide mathematical analyses of the problem.

Assumption 1. ***Linearity*** *[20]*
For every $\mathbf{c}, \mathbf{v} \in \mathbb{R}^n$*, we have:*

$$f(\mathbf{c}, \mathbf{v}) = \mathbf{c}^\top \mathbf{v} \ . \tag{3.1}$$

This linearity assumption entails that the human preference score can be interpreted as a weighted sum of categorical attributes, wherein the vector $\mathbf{v}$ encodes the corresponding weight coefficients. This assumption is justified by the fact that features can be extracted from contexts in order to make the preference function approximately linear.

Assumption 2. ***Hypersphere Assumption***
Let $\mathbf{v}^* \in \mathbb{R}^n$*, we have:*

$$\|\mathbf{v}^*\| = 1 \ . \tag{3.2}$$

This assumption is justified by the fact that, within the given setting, all points that are scalar multiples of the same unit vector have equivalent solutions. In other words, they share the same proportional relationships among the dimensions under consideration.

Assumption 3. ***Preference for Each Sample Assumption*** *[41]*
Let $\mathbf{v}^* \in \mathbb{R}^n$*, let* $\mathbf{c}_1 \sim \mathcal{G}^n$ *and let* $\mathbf{c}_2 \sim \mathcal{Z}^n$*, then:*

$$\forall i \in \{1, \ldots, k\}, \quad \mathbb{P}_{\mathcal{G},\mathcal{Z}}((\mathbf{c}_1^i - \mathbf{c}_2^i)^\top \mathbf{v}^* \neq 0) = 1 \ . \tag{3.3}$$

In preference-based learning framework it is crucial that we can always distinguish between options. This assumption ensures that we never encounter situations where where two samples are indistinguishable in terms of their alignment with $\mathbf{v}^*$.

Assumption 4. ***Independence of the elements of the feature vectors*** *For every of contexts* $\mathbf{c}_1 \sim \mathcal{G}^n, \mathbf{c}_2 \sim \mathcal{Z}^n$*, we have:*

$$\forall i, j \in \{1, \ldots, n\}, \quad i \neq j \implies c_{1,i} \perp c_{1,j} \quad \textit{and} \quad c_{2,i} \perp c_{2,j} \tag{3.4}$$

This assumption simplifies both the modeling process and the computational complexity of the problem and it allows us to study contexts individually.

4 | Proposed Algorithm and Analysis

Following the theoretical foundation, in this chapter we first introduce an algorithm to estimate human preference vectors based on observed data. The algorithm constructs a feasibility region for the preference vector, computes its mean, and normalizes it to obtain a final preference estimate. Then, we use this estimate to predict preferences for new responses by evaluating their alignment with the learned preference structure.

Finally, we present a sample complexity analysis, determining the number of comparisons required to achieve a reliable preference estimate. We analyse the sample complexity under different probabilistic assumptions, including isotropic Gaussian, non-isotropic Gaussian, and discrete feature distributions. We explore different probabilistic models to characterize these feature distributions. By leveraging concentration inequalities, theoretical guarantees on sample complexity are derived, providing insights into the efficiency and reliability of the proposed preference-learning approach.

4.1. Proposed Algorithm

Let $\mathbb{D}$ be a dataset composed of k triplets $(\mathbf{c}_1, \mathbf{c}_2, p)$, where $\mathbf{c}_1$ and $\mathbf{c}_2$ are context representations of different answers, and p denotes the preference label. We now present the Cutting Plane Preference Learner (CPPL) algorithm that leverages the feasibility region derived from the dataset to construct a preference estimation model and make informed decisions on new samples.

Algorithm 4.1 Cutting Plane Preference Learner

```
1:  Collect a dataset $\mathbb{D}$: $\mathbb{D} \leftarrow \{(\mathbf{c}_1^i, \mathbf{c}_2^i, p_i)\}_{i=1}^k$
2:  Find the feasibility region $\mathbf{V}$ defined by the dataset $\mathbb{D}$: find_FP($\mathbb{D}$) $\rightarrow \mathbf{V}$
3:  Compute the mean of $\mathbf{V}$: mean($\mathbf{V}$) $\rightarrow \mathbf{v}$
4:  Normalize $\mathbf{v}$: normalize($\mathbf{v}$) $\rightarrow \hat{\mathbf{v}}$
5:  Receive new sample: $(a_1, \tilde{\mathbf{c}}_1, a_2, \tilde{\mathbf{c}}_2)$
6:  Compute preference prediction:
        $\hat{y}_{\hat{\mathbf{v}}} \leftarrow \begin{cases} 0, & \text{if } (\tilde{\mathbf{c}}_1 - \tilde{\mathbf{c}}_2)^T \hat{\mathbf{v}} > 0 \\ 1, & \text{otherwise} \end{cases}$
7:  if $\hat{y}_{\hat{\mathbf{v}}} = 0$ then
8:      Assign $A \leftarrow a_1$
9:  else
10:     Assign $A \leftarrow a_2$
11: end if
```

The first step of the algorithm involves initializing the dataset $\mathbb{D}$, which comprises k pairs of context vectors $(\mathbf{c}_1, \mathbf{c}_2)$ along with the associated preference p. The preference variable p denotes which of the two context vectors was selected in the observed decision-making process.

Next, the algorithm solves the feasibility problem `find_FP`($\mathbb{D}$) to determine a set of possible preference vectors $\mathbf{V}$. In line 3, they are averaged to derive a representative preference vector $\mathbf{v}$. The choice of the midpoint, defined as the mean for each coordinate of all feasible vectors extracted from $\mathbf{V}$, minimizes the worst-case error that arises when the true preference vector $\mathbf{v}^*$ resides at the extremes of the identified region.

In line 4, the preference vector $\mathbf{v}$ is normalized to ensure it has unit length. This normalization step is crucial for maintaining consistency in comparative computations.

At this stage, the algorithm is equipped with an estimated preference vector $\hat{\mathbf{v}}$, which serves as the foundation for predicting the user's choice. Specifically, the algorithm is designed to identify and suggest the response most aligned with the user's preferences. To achieve this, it evaluates a new decision pair:

- a_1 and a_2 represent the newly available answers.
- $\tilde{\mathbf{c}}_1$ and $\tilde{\mathbf{c}}_2$ denote the corresponding context vectors associated with these choices.

In line 6, the algorithm predicts the user's preference by computing the dot product of the difference between the feature vectors and the learned preference vector $\hat{\mathbf{v}}$:

- If the dot product is positive, $\tilde{\mathbf{c}}_1$ is preferred, leading to $\hat{y}_{\hat{\mathbf{v}}} = 0$.
- Otherwise, $\tilde{\mathbf{c}}_2$ is preferred, resulting in $\hat{y}_{\hat{\mathbf{v}}} = 1$.

Finally, the algorithm determines the final selection A based on the predicted preference:

- If $\hat{y}_{\hat{\mathbf{v}}} = 0$, it selects a_1.
- Otherwise, it selects a_2.

Once the feasibility region $\mathbf{V}$ has been sufficiently reduced and the candidate to represent the user's preferences has been chosen, the algorithm for each new sample does not propose the user a choice but directly provides him with the most suitable answer.

4.2. Context Vector Distribution Dissection

The feasibility problem depends strongly on the statistical properties of the feature vectors $\mathbf{c}_1$ and $\mathbf{c}_2$. However, the underlying distribution of these vectors is unknown and inherently dependent on the LLM.

This uncertainty introduces additional complexity to our analysis, necessitating a formal investigation of the associated random variables. Specifically, the distribution of the difference $\mathbf{c}_1 - \mathbf{c}_2$ is of central importance, as it directly impacts the feasibility conditions of the problem. The choice of probabilistic representation for these distributions significantly influences the theoretical guarantees on sample complexity. Accordingly, the following sections develop a framework for characterizing these distributions and parameterizing their interactions.

Understanding the Structure of $\mathbf{c}_1$ and $\mathbf{c}_2$

A prerequisite to analyzing the statistical properties of $\mathbf{c}_1 - \mathbf{c}_2$ is a precise understanding of the individual components $\mathbf{c}_1$ and $\mathbf{c}_2$.

First, considering the **support** of these feature vectors, each element c_{1j} and c_{2j} takes values from a discrete set $\{0, \ldots, d\}$.

Next, examining their **distribution**, we recognize that the feature vectors may either follow identical distributions or differ in their statistical properties. Assuming a shared distribution simplifies analytical treatment, yet allowing for differences is necessary when comparing responses generated by distinct models.

Another critical aspect is the **independence** of vector components. In many practical scenarios, dependencies exist among elements, necessitating more sophisticated statistical models for accurate representation. Formally establishing these characteristics enables the parameterization of the distribution governing the difference vector $\mathbf{c}_1 - \mathbf{c}_2$.

Probabilistic Representations

To derive meaningful insights into sample complexity, we adopt specific probabilistic models for $\mathbf{c}_1$ and $\mathbf{c}_2$. Our study assumes independence between elements within the same vector, as stated in Assumption 4, while allowing for potentially distinct distributions between the two vectors.

A natural approach is to model the components as **binomial random variables**. Specifically, each element c_{1j} or c_{2j} may be treated as a binomially distributed random variable, i.e., $\text{Binomial}(d, m)$ [42], where d denotes the number of trials and m the probability of success in each trial. This interpretation is particularly relevant when features originate from counting processes or categorical assignments with a fixed number of possible outcomes.

Alternatively, a **Gaussian approximation** [43] may be employed for analytical tractability. In this case, each element c_{1j} and c_{2j} is characterized by a mean μ_{1j} and μ_{2j}, along with variances σ^2_{1j} and σ^2_{2j}, estimated from the underlying discrete distributions. By the central limit theorem, when the binomial parameter d is sufficiently large, the binomial distribution is well approximated by a normal distribution [44]. This Gaussian assumption facilitates closed-form solutions in optimization and inference tasks.

By exploring these different probabilistic models, we enhance our flexibility in characterizing the distribution of $\mathbf{c}_1 - \mathbf{c}_2$. This, in turn, plays a fundamental role in deriving theoretical guarantees on sample complexity and ensuring the robustness of the proposed learning algorithm.

4.3. Sample Complexity

The analysis is divided into three cases. To ensure clarity, each scenario is addressed separately, starting with the simplest case—where $\mathbf{c}_1 - \mathbf{c}_2$ is modeled as an isotropic multivariate Gaussian distribution. The other two cases are treated subsequently and are those of a non-isotropic Gaussian, in which importance sampling will be used to return to the base case and that of a discrete distribution which will require more refined mathematical tools.

To establish an upper bound on the number of samples k required to achieve a small error, we use the feasibility region $\mathbf{V}$ obtained by solving the offline feasibility problem as a measure of error. This region represents the set of feasible solutions within which the true preference vector $\mathbf{v}^*$ is expected to lie.

By quantifying how $\mathbf{V}$ shrinks as the number of samples increases, we can determine the conditions under which the estimated preference vector $\hat{\mathbf{v}}$ remains sufficiently close to $\mathbf{v}^*$.

Definition 1 (Distance $\hat{d}$). *The Euclidean norm between the optimal solution* $\mathbf{v}^*$ *and its corresponding estimator* $\hat{\mathbf{v}}$ *is defined as the distance* $\hat{d}$*:*

$$\hat{d} = \|\hat{\mathbf{v}} - \mathbf{v}^*\| \ . \tag{4.1}$$

Accordingly, our objective is to ensure that the probability $\mathbb{P}(\hat{d} < \epsilon)$ exceeds $1 - \delta$, i.e.,

$$\mathbb{P}(\hat{d} < \epsilon) > 1 - \delta \ . \tag{4.2}$$

In other words, we seek to guarantee, with high probability, that the error associated with $\hat{d}$ remains below a predefined threshold ϵ.

4.3.1. Isotropic Gaussian Distribution Case

Gaussian Approximation

We assume that $\mathbf{c}_1$ follows a multivariate normal distribution with mean μ_1 and diagonal covariance matrix $\mathbf{\Sigma}_1$, ensuring independence among its components. Similarly, $\mathbf{c}_2$ is modeled as an independent multivariate normal vector with mean μ_2 and covariance matrix $\mathbf{\Sigma}_2$, allowing for distinct distributions between the two feature sets.

Since the sum or difference of Gaussian random variables remains Gaussian, the difference $\mathbf{c}_1 - \mathbf{c}_2$ follows a multivariate normal distribution. Applying the linearity of expectation, the mean of the difference vector is given by:

$$\mathbb{E}[\mathbf{c}_1 - \mathbf{c}_2] = \mu_1 - \mu_2 \ .$$

Likewise, assuming independence, the covariance structure satisfies:

$$\mathrm{Cov}[\mathbf{c}_1 - \mathbf{c}_2] = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2 \ .$$

Thus, we conclude the parametrization:

$$\mathbf{c}_1 - \mathbf{c}_2 \sim \mathcal{N}(\mu_1 - \mu_2, \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2) \ .$$

A special case arises when both $\mathbf{c}_1$ and $\mathbf{c}_2$ originate from the same distribution, characterized by identical mean μ and covariance matrix $\mathbf{\Sigma}$. Under this assumption, the expectation simplifies to:

$$\mathbb{E}[\mathbf{c}_1 - \mathbf{c}_2] = \mathbf{0} \ ,$$

indicating a zero-centered difference vector. The covariance structure also simplifies as:

$$\mathrm{Cov}[\mathbf{c}_1 - \mathbf{c}_2] = 2\mathbf{\Sigma} \ .$$

As a result, the difference vector follows:

$$\mathbf{c}_1 - \mathbf{c}_2 \sim \mathcal{N}(\mathbf{0}, 2\mathbf{\Sigma}) \ .$$

This formulation highlights the symmetric, zero-centered nature of the difference distribution and serves as a foundation for further analysis of sample complexity.

2d Case

We initially restrict our consideration to the two-dimensional (2d) scenario to analyze this case study. This simplification facilitates the visualization of the problem and provides a conceptual foundation for understanding the intuition behind the derived sample complexity.

The distance $\hat{d}$ is a chord of the unitary circumference which subtends an angle θ at the circle's center. The interval $\hat{d}$ is determined by the two constraints closest to $\mathbf{v}^*$, which are generated by the distribution of $\mathbf{c}_1 - \mathbf{c}_2$. Imposing this condition is equivalent to ensuring $\hat{\mathbf{v}}$ close to $\mathbf{v}^*$.

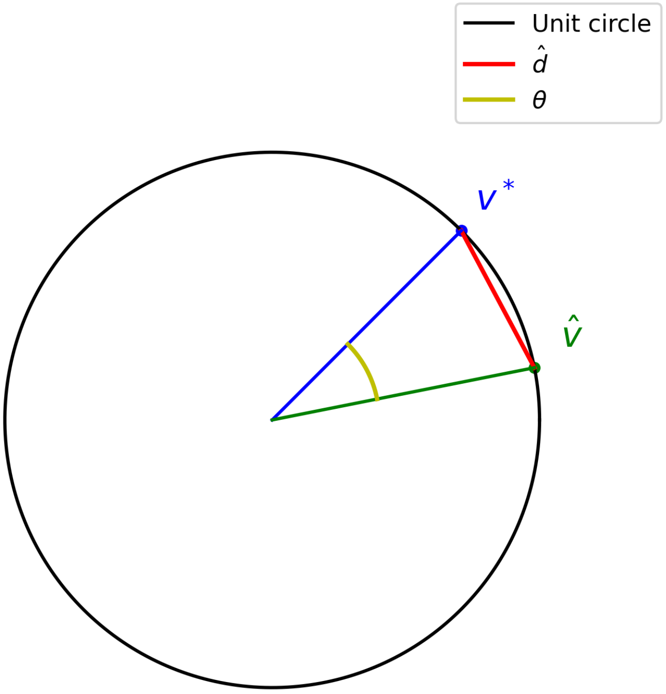


Figure 4.1: 2d sample complexity setting visualization.

Lemma 1. *In a 2d setting* $\hat{d} = \|\hat{\mathbf{v}} - \mathbf{v}^*\|$ *be a chord of the unit circle, and let* θ *be the corresponding central angle. Then:*

$$\|\hat{\mathbf{v}} - \mathbf{v}^*\| \leq \theta \ . \tag{4.3}$$

Proof. The relationship between a chord $\hat{d}$ and its corresponding central angle θ is given by:

$$\hat{d} = 2\sin\left(\frac{\theta}{2}\right) \leq \theta \ .$$

This follows from the well-known inequality $\sin x \leq x$ for $x \geq 0$. □

Now that we have established the relationship between $\hat{d}$ and θ, we proceed to determine the sample complexity. To achieve this, we must first demonstrate that the angle at which the constraints arise is uniformly distributed over the interval $[-\pi, \pi]$.

To establish this result, we analyze the arctangent of the ratio of the distribution $\mathbf{c}_1 - \mathbf{c}_2$.

Lemma 2. *[45] If* X, Y *are independent standard normal random variables and* $\gamma = \arctan 2(X, Y)$*, then* γ *is uniformly distributed over* $[-\pi, \pi]$.

Proof. By hypothesis, X and Y are i.i.d. $\mathcal{N}(0,1)$. Their joint density is

$$f_{X,Y}(x,y) \;=\; \frac{1}{2\pi}\,\exp\!\Big(-\tfrac{x^2+y^2}{2}\Big) \quad \text{for all } (x,y)\in\mathbb{R}^2\;.$$

We define the transformation to polar coordinates by

$$\begin{cases} r \;=\; \sqrt{x^2+y^2}\;, \\ \beta \;=\; \arctan 2(x,y) \;\in\; [-\pi,\pi]\;. \end{cases}$$

We denote this angle by γ in the statement of the theorem, so $\beta=\gamma$.

Fix $r\geq 0$ and $\alpha\in[-\pi,\pi]$. We compute

$$\mathbb{P}[R\leq r,\;\gamma\leq\alpha] \;=\; \mathbb{P}\Big[(X,Y)\in\Omega_{r,\alpha}\Big]\;,$$

where

$$\Omega_{r,\alpha} \;=\; \{(x,y)\in\mathbb{R}^2 : \sqrt{x^2+y^2}\leq r,\; \arctan 2(x,y)\leq\alpha\}\;.$$

Because X and Y are jointly continuous with the given density, we can express this probability as an integral over $\Omega_{r,\alpha}$:

$$\mathbb{P}[R\leq r,\;\gamma\leq\alpha] \;=\; \int_{\Omega_{r,\alpha}} \frac{1}{2\pi}\,\exp\!\Big(-\frac{x^2+y^2}{2}\Big)\,dx\,dy\;.$$

In polar coordinates,

$$x=\rho\,\cos\beta\;,\quad y=\rho\,\sin\beta\;,\quad \text{with Jacobian } \Big|\tfrac{\partial(x,y)}{\partial(\rho,\beta)}\Big| = \rho\;.$$

Thus,

$$dx\,dy \;=\; \rho\,d\rho\,d\beta\;.$$

The region $\Omega_{r,\alpha}$ in polar coordinates becomes

$$\big\{(\rho,\beta) : 0\leq\rho\leq r,\; -\pi\leq\beta\leq\alpha\big\} \quad (\text{assuming } \alpha\geq-\pi)\;.$$

Hence,

$$\mathbb{P}[R\leq r,\;\gamma\leq\alpha] \;=\; \int_{\beta=-\pi}^{\alpha}\int_{\rho=0}^{r} \frac{1}{2\pi}\,\exp\!\Big(-\tfrac{\rho^2}{2}\Big)\,\rho\;d\rho\,d\beta\;.$$

First, integrate with respect to ρ:

$$\int_0^r \rho \, \exp\!\Big(-\tfrac{\rho^2}{2}\Big)\, d\rho \;=\; \Big[-\exp\!\Big(-\tfrac{\rho^2}{2}\Big)\Big]_0^r \;=\; 1-\exp\!\Big(-\tfrac{r^2}{2}\Big)\;.$$

Next, integrate with respect to β from $-\pi$ to α. Since the integrand no longer depends on β,

$$\int_{-\pi}^{\alpha} \frac{1}{2\pi}\, d\beta \;=\; \frac{\alpha-(-\pi)}{2\pi} \;=\; \frac{\alpha+\pi}{2\pi}\;.$$

Putting these pieces together yields

$$\mathbb{P}[R \le r,\ \gamma \le \alpha] \;=\; \left(\frac{\alpha+\pi}{2\pi}\right)\Big(1-\exp\!\Big(-\tfrac{r^2}{2}\Big)\Big)\;.$$

Differentiate the function $\mathbb{P}[R \le r,\ \gamma \le \alpha]$ with respect to r and α. Because

$$\frac{\partial}{\partial r}\big(1-e^{-r^2/2}\big) = r\,e^{-r^2/2}\;, \quad \text{and} \quad \frac{\partial}{\partial \alpha}\big(\tfrac{\alpha+\pi}{2\pi}\big) = \frac{1}{2\pi}\;,$$

we get the joint density of (R,γ):

$$f_{R,\gamma}(r,\alpha) \;=\; \frac{r}{2\pi}\,\exp\!\Big(-\tfrac{r^2}{2}\Big)\,\mathbb{1}_{\{r\ge 0\}}\,\mathbb{1}_{\{-\pi\le\alpha\le\pi\}}\;.$$

To find the density of γ, integrate out r:

$$f_\gamma(\alpha) \;=\; \int_0^\infty \frac{r}{2\pi}\,\exp\!\Big(-\tfrac{r^2}{2}\Big)\,dr\;\mathbb{1}_{\{-\pi\le\alpha\le\pi\}} \;=\; \frac{1}{2\pi}\,\underbrace{\Big[\int_0^\infty r\,\exp\!\Big(-\tfrac{r^2}{2}\Big)\,dr\Big]}_{=1}\;\mathbb{1}_{\{-\pi\le\alpha\le\pi\}}\;.$$

Here we used again the fact that $\int_0^\infty r e^{-r^2/2}\,dr = 1$. Thus

$$f_\gamma(\alpha) \;=\; \frac{1}{2\pi}\,\mathbb{1}_{\{-\pi\le\alpha\le\pi\}}\;,$$

which is precisely the density of the uniform distribution on $[-\pi,\pi]$. Hence $\gamma \sim \mathcal{U}(-\pi,\pi)$, where $\mathcal{U}$ indicates uniform distribution.

□

Corollary 1. *If the components of the vector* $\mathbf{c}_1-\mathbf{c}_2 \in \mathbb{R}^2$ *are independent and isotropic Gaussian, then:*

$$\arctan\left(\frac{(c_1-c_2)_2}{(c_1-c_2)_1}\right) \sim \mathcal{U}[-\pi,\pi]\;. \tag{4.4}$$

Having established that the constraint of our problem in this case are generated uniformly in the plane, we are now prepared to derive the sample complexity, denoted as k.

Theorem 4.1 (2d Isotropic Gaussian Sample Complexity). *Let $\theta \in [0, \pi]$, let $\delta \in \left[0, \frac{1}{2}\right]$, let $\mathcal{U}_i \sim \mathcal{U}[-\pi, \pi]$ independent for $i \in [k]$. Then, to achieve the condition*

$$1 - \delta \leq \mathbb{P}\left(\exists i : \mathcal{U}_i \in \left[-\frac{\theta}{2}, 0\right] \cap \exists j : \mathcal{U}_j \in \left[0, \frac{\theta}{2}\right]\right) = \mathbb{P}_{2D} \ , \tag{4.5}$$

the number of required samples satisfies:

$$k \geq 4\pi \cdot \frac{\ln(2\delta^{-1})}{\theta} \ . \tag{4.6}$$

Proof. Calculate the probability of constraints in the intervals:

$$\mathbb{P}(\exists\, i \in \{1, \ldots, k\} : \mathcal{U}_i \in [-\theta/2, 0] \quad \wedge \quad \exists\, j \in \{1, \ldots, k\} : \mathcal{U}_j \in [0, \theta/2]) \ .$$

Use independence of events to combine the probabilities:

$$\geq \mathbb{P}(\exists\, i \in \{1, ..., k\} : \mathcal{U}_i \in [-\theta/2, 0]) \cdot \mathbb{P}(\exists\, j \in \{1, ..., k\} : \mathcal{U}_j \in [0, \theta/2]) \ .$$

This becomes:

$$= (1 - \mathbb{P}(\forall\, i \in \{1, \ldots, k\} : \mathcal{U}_i \notin [-\theta/2, 0])) \cdot (1 - \mathbb{P}(\forall\, j \in \{1, \ldots, k\} : \mathcal{U}_j \notin [0, \theta/2])) \ .$$

The individual probabilities are calculated based on the proportion of the angle θ to the full circle 2π using the independence of the $\mathcal{U}_i \in [k]$. The expression:

$$\left(1 - \left(\frac{2\pi - \theta/2}{2\pi}\right)^k\right)^2 ,$$

reflects the likelihood that at least one constraint falls within each interval.

Expand the squared term:

$$= 1 + \left(\frac{2\pi - \theta/2}{2\pi}\right)^{2k} - 2\left(\frac{2\pi - \theta/2}{2\pi}\right)^k \geq 1 - 2\left(\frac{2\pi - \theta/2}{2\pi}\right)^k \ .$$

Setting the latter $\geq 1-\delta$, rearrange terms to form an inequality involving δ:

$$
\begin{aligned}
\delta &\geq 2\left(\frac{2\pi-\theta/2}{2\pi}\right)^{k} ,\\
\log\left(\frac{2}{\delta}\right) &\leq k\log\left(\frac{2\pi}{2\pi-\theta/2}\right) ,\\
k &\geq \log\left(\frac{2}{\delta}\right)\frac{1}{\log\left(\frac{1}{1-\frac{\theta}{4\pi}}\right)} .
\end{aligned}
$$

Upper bound using the inequality $\frac{1}{\log\left(\frac{1}{1-x}\right)} \leq \frac{1}{x}$, with $x \in (0,1)$:

$$k \geq \log\left(\frac{2}{\delta}\right)\cdot\frac{4\pi}{\theta} .$$

□

Generalization

To extend sample complexity analysis from the two-dimensional case to an arbitrary n-dimensional setting, we begin by examining the 2-norm of the difference between $\mathbf{v}^*$ and $\hat{\mathbf{v}}$ in higher dimensions. In an n-dimensional space, we can divide this 2-norm using $\lceil\frac{n}{2}\rceil$ different angular parameters $(\theta_1, \theta_2, \ldots, \theta_{\lceil\frac{n}{2}\rceil})$, where each of these angles determines the orientation of constraints. Moreover, as established in Lemma 2, each angular parameter θ_i is independently and uniformly distributed over the interval $[-\pi, \pi]$, which forms the basis for our probabilistic analysis.

As done in the 2d case, we set the amplitude for each angle to θ. Geometrically, this condition ensures that the feasible configuration is constrained within a hypercube whose volume we want proportional to ϵ.

Theorem 4.2 (Sample Complexity Isotropic Gaussian in n-Dimensions)**.** *In an n-dimensional space, the number of required samples k to satisfies $\mathbb{P}(\hat{d}<\epsilon) > 1-\delta$ is:*

$$k \geq \sqrt{\left\lceil\frac{n}{2}\right\rceil}\ \log\left(\frac{2\left\lceil\frac{n}{2}\right\rceil}{\delta}\right)\frac{4\pi}{\epsilon}. \tag{4.7}$$

Proof.

$$\begin{aligned} \|\mathbf{v}^* - \hat{\mathbf{v}}\|_2^2 \;&=\; \sum_{i=1}^{n} \left|v_i^* - \hat{v}_i\right|^2 \\ &\leq \begin{cases} \sum_{j=1}^{\frac{n}{2}} \left(\left|v_{2j-1}^* - \hat{v}_{2j-1}\right|^2 + \left|v_{2j}^* - \hat{v}_{2j}\right|^2\right), & \text{if } n \text{ even} \\ \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left(\left|v_{2j-1}^* - \hat{v}_{2j-1}\right|^2 + \left|v_{2j}^* - \hat{v}_{2j}\right|^2\right) \;+\; \left|v_n^* - \hat{v}_n\right|^2 + \left|v_1^* - \hat{v}_1\right|^2 \,, & \text{otherwise} \end{cases} \,, \end{aligned}$$

in the case of an odd number of n, we establish an upper bound of the squared norm by adding an additional term to the summation, where we pair the n-th coordinate with an arbitrarily chosen one, specifically the first. Using the result of Lemma 1 and the independence:

$$\leq \begin{cases} \sum_{j=1}^{\frac{n}{2}} \left(\;2\sin\frac{\theta_j}{2}\right)^2, & \text{if } n \text{ even} \\ \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left(2\sin\frac{\theta_j}{2}\right)^2 \;+\; \left(2\sin\frac{\theta_{\lfloor \frac{n}{2} \rfloor+1}}{2}\right)^2 \,, & \text{otherwise} \end{cases} .$$

Since we set all θ_j equal to the same θ:

$$\leq \left\lceil \frac{n}{2} \right\rceil \, \epsilon^2 \;\longrightarrow\; \|\mathbf{v}^* - \hat{\mathbf{v}}\|_2 \;\leq\; \sqrt{\left\lceil \frac{n}{2} \right\rceil}\, \epsilon \;.$$

To conclude, we observe that this setting ensures:

$$\begin{aligned} &\theta_j \;\leq\; \varepsilon \;\Longrightarrow\; \log\!\left(\frac{2}{\delta}\right)\frac{4\pi}{\epsilon} \;,\; \text{w.p. } \; 1-\delta \;, \\ &\theta_1, \ldots, \theta_{\lceil \frac{n}{2} \rceil} \;\leq\; \varepsilon \;\Longrightarrow\; \log\left(\frac{2\left\lceil \frac{n}{2} \right\rceil}{\delta}\right)\frac{4\pi}{\epsilon} \;,\; \text{w.p. } \; 1-\delta \;, \\ &\|\mathbf{v}^* - \mathbf{v}^*\|_2 \;\leq\; \varepsilon \;\Longrightarrow\; \sqrt{\left\lceil \frac{n}{2} \right\rceil}\,\log\left(\frac{2\left\lceil \frac{n}{2} \right\rceil}{\delta}\right)\frac{4\pi}{\epsilon} \;,\; \text{w.p. } \; 1-\delta \;. \end{aligned}$$

□

This result demonstrates that the sample complexity scales as square root in the dimensionality $n-1$. This reflects the intrinsic geometric complexity of higher dimensional spaces: as the number of angular parameters increases, the difficulty of obtaining high-probability configurations increases.

Intuition Behind the 3D Case

If we apply what has been said in the three-dimensional case, the required solution corresponds to the area of a square projected onto the surface of the unit sphere.

This characterization is an over-approximation of the current feasibility area $\mathbf{V}$: determining a sample complexity with high probability for this area guarantees the same guarantee for $\mathbf{V}$. This allows us to simplify the analysis while obtaining robustness in varying all possible forms of $\mathbf{V}$.

The error caused by the over-approximation is bounded by the diagonal of the projected square that we set as ϵ and represents the maximum possible distance between any two points within the square.

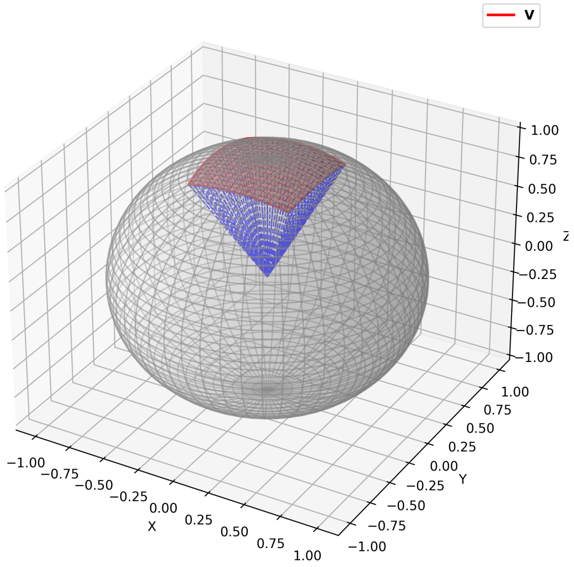


Figure 4.2: 3d sample complexity setting visualization.

4.3.2. Non-Isotropic Gaussian Case

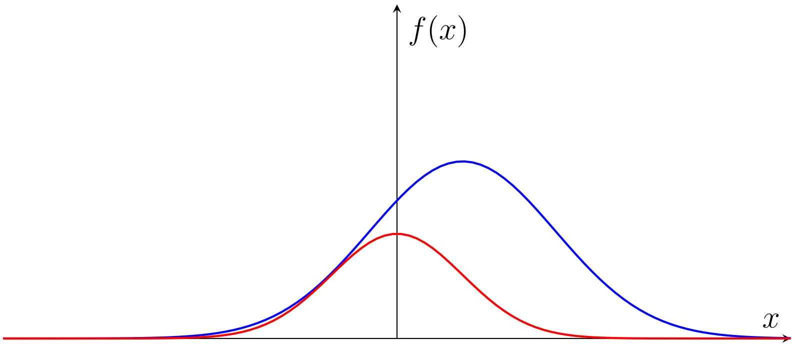


Figure 4.3: Example of a non-isotropic Gaussian G (in blue) that dominates another Gaussian Z (in red).

Consider a non-isotropic Gaussian distribution $G \sim \mathcal{N}(\mu, \Sigma)$, where the mean deviates from zero, thereby introducing a bias in the generation of constraints. To analyze this case, we select an isotropic Gaussian $Z \sim \mathcal{N}(0, \sigma^2 I)$ that is fully dominated by G, that is $\forall x \in \mathbb{R}, f(x)_G \geq f(x)_Z$ (as depicted in Figure 4.3).

Lemma 3 (Importance Sampling Bound)**.** *The probability under Z can be rewritten using importance sampling as:*

$$\mathbb{P}_Z(E) = \left(\mathbb{E}_G\left[\left(\frac{z(x)}{g(x)}\right)^{\alpha}\right]\right)^{1/\alpha} \cdot \left(\mathbb{P}_G(E)\right)^{1-1/\alpha} \quad . \tag{4.8}$$

Proof. Applying Hölder's inequality to the expectation term gives the desired result.

$$\begin{aligned}
\mathbb{P}_Z(E) &= \mathbb{E}_Z[\mathbb{1}_{\{E\}}] \\
&= \mathbb{E}_G\left[\frac{z(x)}{g(x)}\mathbb{1}_{\{E\}}\right] \\
&\leq \left(\mathbb{E}_G\left[\left(\frac{z(x)}{g(x)}\right)^{\alpha}\right]\right)^{1/\alpha} \cdot \left(\mathbb{E}_G[\mathbb{1}_{\{E\}}^{\frac{\alpha}{\alpha-1}}]\right)^{1-1/\alpha} \\
&= \left(\mathbb{E}_G\left[\left(\frac{z(x)}{g(x)}\right)^{\alpha}\right]\right)^{1/\alpha} \cdot \left(\mathbb{P}_G(E)\right)^{1-1/\alpha} \quad .
\end{aligned}$$

Where:

- $\mathbb{P}_Z(E)$: The probability of event E under the probability measure associated with Z.
- $\mathbb{E}_Z[\cdot]$: Expectation taken concerning the probability distribution of Z.
- $\mathbb{1}_{\{E\}}$: Indicator function of the event E, which takes the value 1 if $x \in E$ and 0 otherwise.
- $\mathbb{E}_G[\cdot]$: Expectation taken concerning the probability distribution of G.
- $z(x)$: Probability density function (PDF) of the distribution associated with Z.
- $g(x)$: Probability density function (PDF) of the distribution associated with G.
- $\frac{z(x)}{g(x)}$: Likelihood ratio that reweights probabilities under G to probabilities under Z.
- α: A parameter ≥ 1 , related to Hölder's inequality, which is used to bound expectations.

□

As a consequence, we can express:

$$\mathbb{P}_G(E) \geq \frac{\mathbb{P}_Z(E)^{\alpha/(\alpha-1)}}{\left(\int z(x)^\alpha g(x)^{1-\alpha}\,dx\right)^{1/(\alpha-1)}}\,. \tag{4.9}$$

Having determined this, we have found an additive term that adds to the sample complexity of the isotropic Gaussian case.

4.3.3. Discrete Case

If G is a discrete-valued distribution, the previous integral is no longer well-defined. In such cases, we approximate G with a Gaussian distribution, which introduces an irreducible error term. This error impacts the sample complexity by imposing a lower bound on the angle θ, preventing it from being reduced arbitrarily. This constraint arises because there exists a region $\mathbf{V}^*$ in which all points represent equivalent solutions to our problem. We now introduce our approach to derive the sample complexity in the discrete case.

Lemma 4 (Error Propagation)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Let $\boldsymbol{X} \sim \mathcal{P}^n$ and $\boldsymbol{Y} \sim \mathcal{Q}^n$ be real-valued component-wise independent random vectors, with $\mathcal{P}$ and $\mathcal{Q}$ being two generic n-dimensional distribution. Assume that there exists a set of* *deterministic bounds* $\{\Delta_i\}_{i\in[n]} \in [0,1]^n$ *such that*

$$\forall i \in [n] \quad \sup_{t\in\mathbb{R}} |\mathbb{P}(X_i \leq t) - \mathbb{P}(Y_i \leq t)| \leq \Delta_i \ . \tag{4.10}$$

Then we have

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(f(\boldsymbol{X}) \leq t) - \mathbb{P}(f(\boldsymbol{Y}) \leq t)| \leq ||\nabla f||_\infty \sum_{i\in[n]} \Delta_i \ , \tag{4.11}$$

where $||\nabla f||_\infty$ *indicates the infinity norm of the gradient of* f*.*

Proof. The proof is a consequence of the mean-value theorem for integrals [46] and Hölder inequality [47]. □

Corollary 2 (Error Propagation for Normal Approximation of Binomials)**.** *Let* $X_i \sim \mathcal{P}^n = \{Bin(d_i, m_i)\}_{i\in[n]}$ *and* $Y_i \sim \mathcal{Q}^n = \{\mathcal{N}(d_i m_i, d_i m_i(1-m_i)\}_{i\in[n]}$*. Then, we have [48][49]*

$$\forall i \in [n] \quad \sup_{t\in\mathbb{R}} |\mathbb{P}(X_i \leq t) - \mathbb{P}(Y_i \leq t)| \leq \frac{1}{\sqrt{d_i m_i (1-m_i)}} \ , \tag{4.12}$$

and, subsequently by Lemma 4, assuming that $m_i = m$ *and* $d_i = d$ *for every* $i \in [n]$*, we have*

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(f(\boldsymbol{X}) \leq t) - \mathbb{P}(f(\boldsymbol{Y}) \leq t)| \leq ||\nabla f||_\infty \frac{n}{\sqrt{dm(1-m)}} \ . \tag{4.13}$$

Proof. Equation (4.12) is a consequence of the Berry-Esseen Theorem for the normal approximation of random variables. Equation (4.13) is a trivial consequence of Lemma 4, Equation (4.12) and the fact that all random variables are assumed to be iid. □

As seen previously, the constraints are determined by the arctangent of the ratio of our distributions, also in this case we study it.

Lemma 5 (Gradient of Arctan Function)**.** *Let* $f(x_1, x_2, x_3, x_4) := \arctan\left(\frac{x_1 - x_2}{x_3 - x_4}\right)$*, for* $\boldsymbol{x} := (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$*. Then*

$$||\nabla f|| = \frac{1}{1+r^2} \sqrt{2\frac{r^2}{(x_3-x_4)^2} + 2\frac{1}{(x_3-x_4)^2}} \ , \tag{4.14}$$

where $r^2 := \left(\frac{x_1-x_2}{x_3-x_4}\right)^2$. *In particular, assuming* $x \in \mathbb{N}^4$ *and* $x_3 \neq x_4$*, it holds that:*

$$||\nabla f||_\infty \leq \sqrt{2}\ . \tag{4.15}$$

Proof. This result is a consequence of the definition of gradient and simple calculations. □

Lemma 6 (Equality of two Binomial)**.** *Let* $X_1, X_2 \overset{iid}{\sim} \text{Binomial}(d, m)$*, then the probability* $I_d := \mathbb{P}\big(X_1 = X_2\big)$ *is defined as:*

$$I_d \ = \sum_{i=0}^{d} \binom{d}{i}^2 m^{2i} \left(1-m\right)^{2(d-i)}\ . \tag{4.16}$$

Proof.

$$\mathbb{P}\big(X_1 = X_2\big) \ = \sum_{i=0}^{d} \mathbb{P}\big(X_1 = i\big)\mathbb{P}\big(X_2 = i\big)\ .$$

The PMF of a Binomial(d, p) random variable X is

$$\mathbb{P}(X = i) \ = \binom{d}{i} m^i \left(1-m\right)^{d-i}\ .$$

Therefore,

$$\mathbb{P}(X_1 = i) \ = \binom{d}{i} m^i \left(1-m\right)^{d-i}, \quad \mathbb{P}(X_2 = i) \ = \binom{d}{i} m^i \left(1-m\right)^{d-i},$$

hence,

$$\mathbb{P}\big(X_1 = X_2\big) \ = \sum_{i=0}^{d} \left[\binom{d}{i} m^i \left(1-m\right)^{d-i}\right] \times \left[\binom{d}{i} m^i \left(1-m\right)^{d-i}\right]\ .$$

Combining similar terms gives

$$\mathbb{P}\big(X_1 = X_2\big) \ = \sum_{i=0}^{d} \binom{d}{i}^2 m^{2i} \left(1-m\right)^{2(d-i)}\ .$$

□

Lemma 7 (I_d Bound)**.** *Let* $X, Y \overset{iid}{\sim} \text{Binomial}(d, m)$*, let* $I_d = \mathbb{P}\big(X = Y\big)$*, then it holds that:*

$$I_d \leq \left(\frac{1}{\sqrt{2}} + \frac{1}{2\sqrt{\pi}}\right) \frac{1}{\sqrt{dm(1-m)}} \ . \tag{4.17}$$

Proof.

$$X, Y \overset{iid}{\sim} \text{Binomial}(d, m), \ X_i, Y_i \overset{iid}{\sim} \text{Bernoulli}(m) \ .$$

We have:

$$X = \sum_{i=1}^{d} X_i, \quad Y = \sum_{i=1}^{d} Y_i \ .$$

So we can define Z:

$$Z := X - Y = \sum_{i=1}^{d} (X_i - Y_i) = \sum_{i=1}^{d} Z_i \ ,$$

where Z_i are iid.

$$\begin{aligned}
&\mathbb{E}[Z_i] = 0 \ , \ \text{Var}(Z_i) = \text{Var}(X_i) + \text{Var}(Y_i) = 2m(1-m) \ , \\
&\mathbb{E}[Z_i^3] = |-1| \cdot \mathbb{P}(X_i = 0, Y_i = 1) + 1 \cdot \mathbb{P}(X_i = 1, Y_i = 0) + 0 \cdot \mathbb{P}(X_i = Y_i) \\
&\ = \mathbb{P}(X_i = 0, Y_i = 1) + \mathbb{P}(X_i = 1, Y_i = 0) = 2m(1-m) \ .
\end{aligned}$$

Applying Berry-Esseen theorem:

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(Z \leq t) - \mathbb{P}(G \leq t)| \leq \frac{C}{\sqrt{d}} \frac{\mathbb{E}[|Z_i|^3]}{\text{Var}(Z_i)^{3/2}} \ ,$$

where $G \sim \mathcal{N}(0, 2md(1-m))$.

Since $C \leq \frac{1}{2}$:

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(Z \leq t) - \mathbb{P}(G \leq t)| \leq \frac{1}{2} \frac{1}{\sqrt{2dm(1-m)}} \ . \tag{4.18}$$

We are to bound:

$$\mathbb{P}(Z = 0) = \mathbb{P}(-1 < Z < 1) = \underbrace{\mathbb{P}(-1 < G < 1)}_{(A)} + \underbrace{\mathbb{P}(-1 < Z < 1) - \mathbb{P}(-1 < G < 1)}_{(B)} \ .$$

Bound of term (A):

$$(A) \leq \int_{-1}^{1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx, \text{with } \sigma^2 = 2dm(1-m) \ ,$$
$$(A) \leq \frac{1}{\sqrt{2\pi\sigma^2}} = \frac{1}{2\sqrt{\pi dm(1-m)}} \ .$$

Bound of term (B):

$$\begin{aligned}(B) &= \mathbb{P}(Z < 1) - \mathbb{P}(Z \leq -1) + \mathbb{P}(G < 1) - \mathbb{P}(G \leq -1) \\ &\leq |\mathbb{P}(Z < 1) - \mathbb{P}(G < 1)| + |\mathbb{P}(Z \leq 1) - \mathbb{P}(G \leq 1)| \ .\end{aligned}$$

Using the Equation 4.18:

$$(B) \leq \frac{1}{\sqrt{2dm(1-m)}} \ .$$

Putting the bounds of (A) and (B) together:

$$(A) + (B) \leq \left(\frac{1}{\sqrt{2}} + \frac{1}{2\sqrt{\pi}}\right) \frac{1}{\sqrt{dm(1-m)}} \ .$$

□

Corollary 3. *Let* $f(\boldsymbol{x}) = \arctan\left(\frac{x_1 - x_2}{x_3 - x_4}\right)$. *Let* $\boldsymbol{X_i} \sim \mathcal{P} = Bin(d, m)$. *Then, with probability* I_d, *we have*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(f(\boldsymbol{X_i}) \leq t) - \mathbb{P}(U \leq t)| \leq \frac{4\sqrt{2}}{\sqrt{dm(1-m)}} \ , \tag{4.19}$$

where $U \sim \mathcal{U}[-\pi, \pi]$ *is a uniform random variable.*

Proof. First, note that the event $X_3 = X_4$ occurs with probability proportional to $\frac{1}{\sqrt{d}}$. Then, given $\boldsymbol{Y} \sim \mathcal{Q}^4 = \mathcal{N}(dm, dm(1-m))$, we have that both $Y_1 - Y_2$ and $Y_3 - Y_4$ are distributed as zero-mean normal random variables. Thus, we have that $f(\boldsymbol{Y}) \sim \mathcal{U}[-\pi, \pi]$, since the arctan of the ratio of zero mean normal random variables is distributed as a uniform in $[-\pi, \pi]$ as demonstrated in Lemma 2. The proof can be completed by simple applications of Corollary 2 and Lemma 5. □

Theorem 4.3 (Sample Complexity for Binomial in 2d)**.** *Let* $\delta \in \left(0, \frac{1}{2}\right)$. *Let* $\{X_{i,j}\}_{i \in [k], j \in [4]} \sim Bin(d, m)^{k \times 4}$ *be a dataset. Then, the expected number of samples* k *required to obtain:*

$$\mathbb{P}\left(\left\{\exists i \in [k] : \widetilde{U}_i \in \left[-\frac{\theta}{2}, 0\right]\right\} \cap \left\{\exists j \in [k] : \widetilde{U}_j \in \left[0, \frac{\theta}{2}\right]\right\}\right) \geq 1 - \delta \ , \tag{4.20}$$

where $\widetilde{U}_i \coloneqq \arctan\left(\frac{X_{i,1}-X_{i,2}}{X_{i,3}-X_{i,4}}\right)$*, is;*

$$k \geq \frac{\log\left(\frac{2}{\delta}\right)}{\log\left(\frac{1}{\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d}\right)},\ \textit{with}\ \frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d \leq 1\ , \tag{4.21}$$

where $\Delta_{d,m} \coloneqq \frac{4\sqrt{2}}{\sqrt{dm(1-m)}}$ *and* $I_d \leq \left(\frac{1}{\sqrt{2}}+\frac{1}{2\sqrt{\pi}}\right)\frac{1}{\sqrt{dm(1-m)}}$.

Proof. We have to lower-bound the LHS:

$$\begin{aligned}
&\mathbb{P}\left(\left\{\exists i \in [k] : \widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]\right\} \cap \left\{\exists j \in [k] : \widetilde{U}_j \in \left[0,\frac{\theta}{2}\right]\right\}\right)\\
&= \mathbb{P}\left(\left\{\exists i \in [k] : \widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]\right\}\right)\mathbb{P}\left(\left\{\exists j \in [k] : \widetilde{U}_j \in \left[0,\frac{\theta}{2}\right]\right\}\right)\\
&= \left(1-\mathbb{P}\left(\left\{\forall i \in [k] : \widetilde{U}_i \notin \left[-\frac{\theta}{2},0\right]\right\}\right)\right)\left(1-\mathbb{P}\left(\left\{\forall j \in [k] : \widetilde{U}_j \notin \left[0,\frac{\theta}{2}\right]\right\}\right)\right)\\
&= \left(1-\prod_{i\in[k]}\mathbb{P}\left(\widetilde{U}_i \notin \left[-\frac{\theta}{2},0\right]\right)\right)\left(1-\prod_{i\in[k]}\mathbb{P}\left(\widetilde{U}_j \notin \left[0,\frac{\theta}{2}\right]\right)\right)\ .
\end{aligned}$$

Now, we noticed that:

$$\begin{aligned}
&\mathbb{P}\left(\widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]\right)\\
&= \mathbb{P}\left(\widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]|X_1 = X_2\right)\mathbb{P}\left(X_1 = X_2\right) + \mathbb{P}\left(\widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]|X_1 \neq X_2\right)\mathbb{P}\left(X_1 \neq X_2\right)\\
&\leq I_d + \mathbb{P}\left(\widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]|X_1 \neq X_2\right)\ .
\end{aligned}$$

Given that :

$$\begin{aligned}
&\mathbb{P}\left(\widetilde{U}_i \in \left[-\frac{\theta}{2},0\right]|X_1 = X_2\right) \leq 1\\
&\mathbb{P}\left(X_1 \neq X_2\right) = 1 - I_d \leq 1\ .
\end{aligned}$$

So, we have:

$$\begin{aligned}
&\left(1-\prod_{i\in[k]}\mathbb{P}\left(\widetilde{U}_i\notin\left[-\frac{\theta}{2},0\right]\right)\right)\left(1-\prod_{i\in[k]}\mathbb{P}\left(\widetilde{U}_j\notin\left[0,\frac{\theta}{2}\right]\right)\right)\\
&\geq\left(1-\prod_{i\in[k]}\left[\mathbb{P}\left(\widetilde{U}_i\leq-\frac{\theta}{2}\right)+1-\mathbb{P}\left(\widetilde{U}_i\leq 0\right)+I_d\right]\right)\cdot\\
&\cdot\left(1-\prod_{i\in[k]}\left[\mathbb{P}\left(\widetilde{U}_i\leq 0\right)+1-\mathbb{P}\left(\widetilde{U}_i\leq\frac{\theta}{2}\right)+I_d\right]\right)
\end{aligned}$$

$$\begin{aligned}
&\geq\left(1-\prod_{i\in[k]}\left[\mathbb{P}\left(U_i\leq-\frac{\theta}{2}\right)+1-\mathbb{P}\left(U_i\leq 0\right)+2\Delta_{d,m}+I_d\right]\right)\cdot\\
&\cdot\left(1-\prod_{i\in[k]}\left[\mathbb{P}\left(U_i\leq 0\right)+1-\mathbb{P}\left(U_i\leq\frac{\theta}{2}\right)+2\Delta_{d,m}+I_d\right]\right)\\
&=\left(1-\prod_{i\in[k]}\left[\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d\right]\right)^2\\
&=\left(1-\left(\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d\right)^k\right)^2\\
&\geq 1-2\left(\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d\right)^k\geq 1-\delta\ .
\end{aligned}$$

The first inequality is a consequence of Corollary 3. The remaining steps are simple calculations. We solve the following inequality in k:

$$\begin{aligned}
\left(\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d\right)^k&\leq\frac{\delta}{2}\ ,\\
k\log\left(\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d\right)&\leq\log\left(\frac{\delta}{2}\right)\ ,\\
\frac{\log\left(\frac{2}{\delta}\right)}{\log\left(\frac{1}{\frac{2\pi-\frac{\theta}{2}}{2\pi}+2\Delta_{d,m}+I_d}\right)}&\leq k\ .
\end{aligned}$$

The proof can be concluded by noting that the true sample complexity also accounts for the discarded samples, i.e., the ones where $X_{i,3} = X_{i,4}$, and that sample complexity is valid only with $\frac{2\pi - \frac{\theta}{2}}{2\pi} + 2\Delta_{d,m} + I_d \leq 1$. □

As done in the isotropic Gaussian setting, we can extend our reasoning to n-dimensions by defining our generalized sample complexity for the Discrete case:

$$k \geq \frac{\sqrt{\left\lceil \frac{n}{2} \right\rceil \log\left(\frac{2\left\lceil \frac{n}{2} \right\rceil}{\delta}\right)}}{\log\left(\frac{1}{\frac{2\pi - \frac{\theta}{2}}{2\pi} + 2\Delta_{d,m} + I_d}\right)} . \tag{4.22}$$

5 | Experiment and Results

In this chapter, we present experimental investigations to validate the theoretical framework we established in previous chapters. We analyze constraint behavior in the feasibility problem, examine feasibility region contraction under different configurations, and evaluate the performance of the proposed preference-learning algorithm. Each experiment follows a structured methodology, incorporating statistical techniques to ensure reliability.

First, we investigate how constraints emerge in the feasibility problem by generating constraint vectors under binomial and Gaussian distributions. We then examine how these regions shrink as the number of constraints increases, computing their average size across multiple runs and exploring different estimators for the optimal solution.

Next, we extend the theoretical analysis of the arctangent of the ratio of realizations from different distributions. While prior work shows a uniform distribution for isotropic Gaussian cases, we explore deviations under non-isotropic and binomial distributions to assess their effect on constraint uniformity.

Beyond synthetic simulations, we evaluate the Preference Dissection dataset [50], which contains AI-generated responses and corresponding evaluations. We preprocess the dataset, extract numerical representations, and conduct a feature selection process to retain the most informative attributes while reducing redundancy.

Finally, we assess the Cutting Plane Preference Learner algorithm by reconstructing preference vectors using both real and synthetic data. We evaluate its predictive performance across multiple fixed reference points, benchmarking it against a majority-voting baseline according to accuracy, precision, and recall.

The remainder of this chapter details each experiment's methodology, results, and key observations on constraint behavior and preference learning.

5.1. Constraint Simulation Experiment

The initial set of experiments conducted as part of our investigation focuses on how constraints emerge within the feasibility problem framework, as outlined in Section 3.2. Specifically, we generated the constraint vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ under the assumption that, in various instances, their elements are sampled from either binomial or Gaussian distributions with predefined parameters. The objective of this approach is to examine how the resulting constraints manifest geometrically, either in a two-dimensional plane or in three-dimensional space. For visualization purposes, we confined our analysis to these two spatial dimensions.

The experimental design incorporates six distinct parameter configurations, which are as follows:

- **Binomial distribution with identical success probability (m) fixed at 0.5**:

$$c_{1j} \sim \text{Binomial}(3, 0.5), \quad c_{2j} \sim \text{Binomial}(3, 0.5).$$

- **Gaussian distribution with identical mean (1.5) and small variance (0.1)**:

$$c_{1j} \sim \mathcal{N}(1.5, 0.1), \quad c_{2j} \sim \mathcal{N}(1.5, 0.1).$$

- **Gaussian distribution with identical mean (1.5) but large variance (3)**:

$$c_{1j} \sim \mathcal{N}(1.5, 3), \quad c_{2j} \sim \mathcal{N}(1.5, 3).$$

- **Gaussian distribution with distinct means and small variance (0.1)**:

$$c_{1j} \sim \mathcal{N}(3, 0.1), \quad c_{2j} \sim \mathcal{N}(0, 0.1).$$

- **Gaussian distribution with distinct means and large variance (3)**:

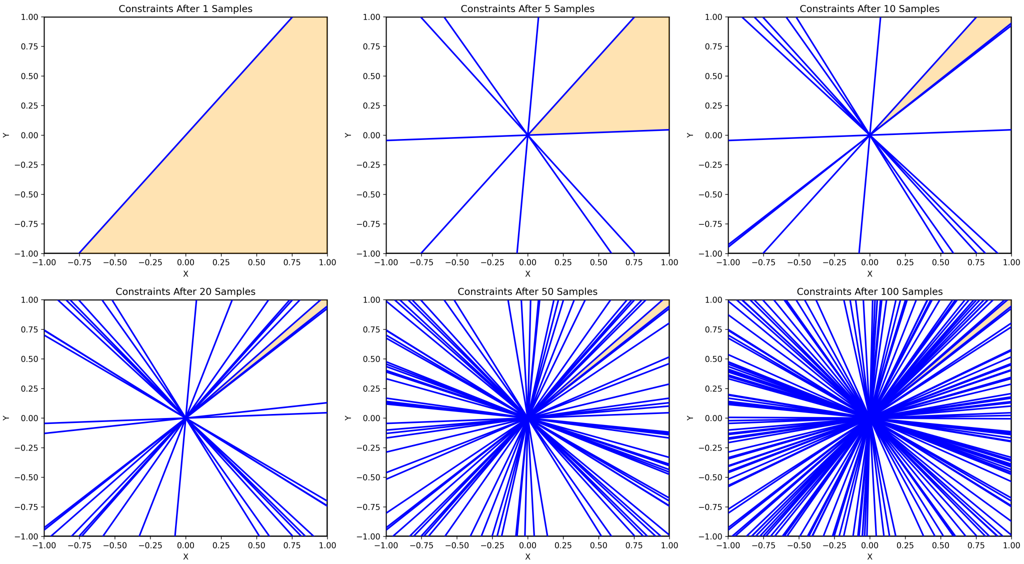
$$c_{1j} \sim \mathcal{N}(3, 3), \quad c_{2j} \sim \mathcal{N}(0, 3).$$

- **Binomial distribution with differing success probabilities**:

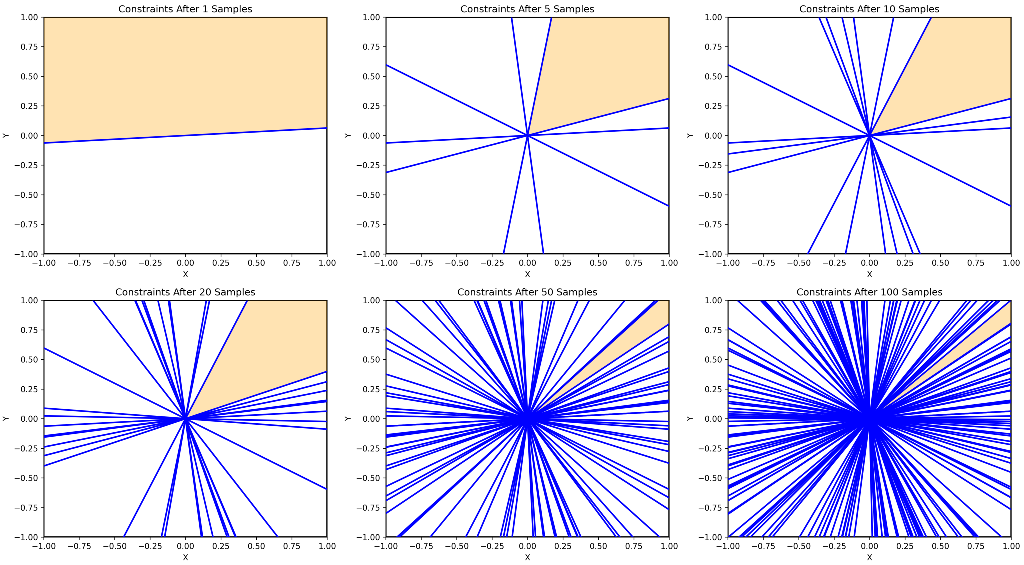
$$c_{1j} \sim \text{Binomial}(3, 0.8), \quad c_{2j} \sim \text{Binomial}(3, 0.2).$$

5.1.1. Visualization of the Constraints

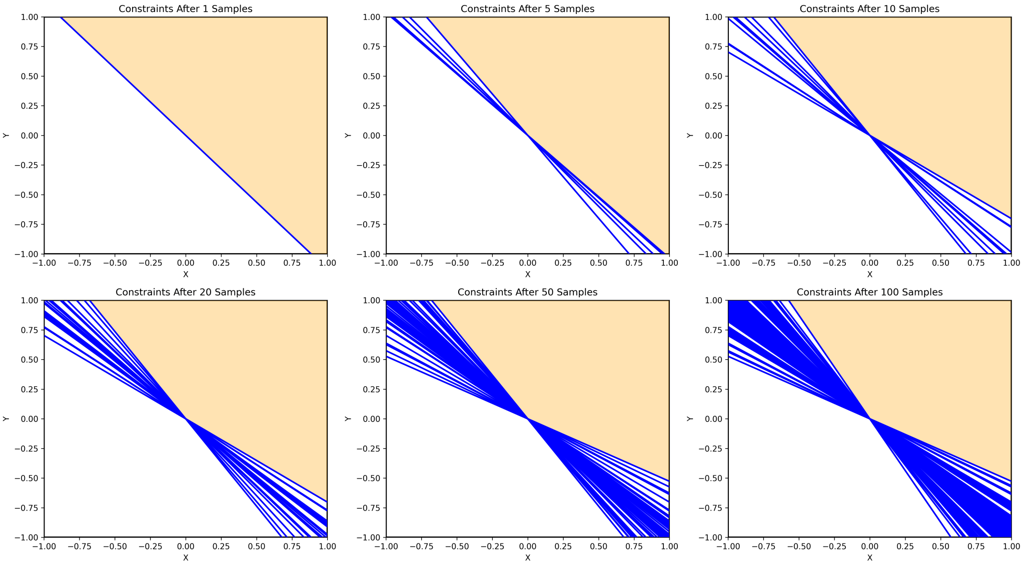
In the initial phase of this experiment, we simulated a run of our problem in two-dimensional space using Gaussian configurations to analyze the arrangement of constraints and the evolution of the feasibility region $\mathbf{V}$ (highlighted in yellow in the images) throughout the experiment. We chose $(1, 1)$ as the fixed point $\mathbf{v}^*$. From Figure 5.1 we can deduce that if the mean is different from 0 a bias is introduced in the generation of constraints.



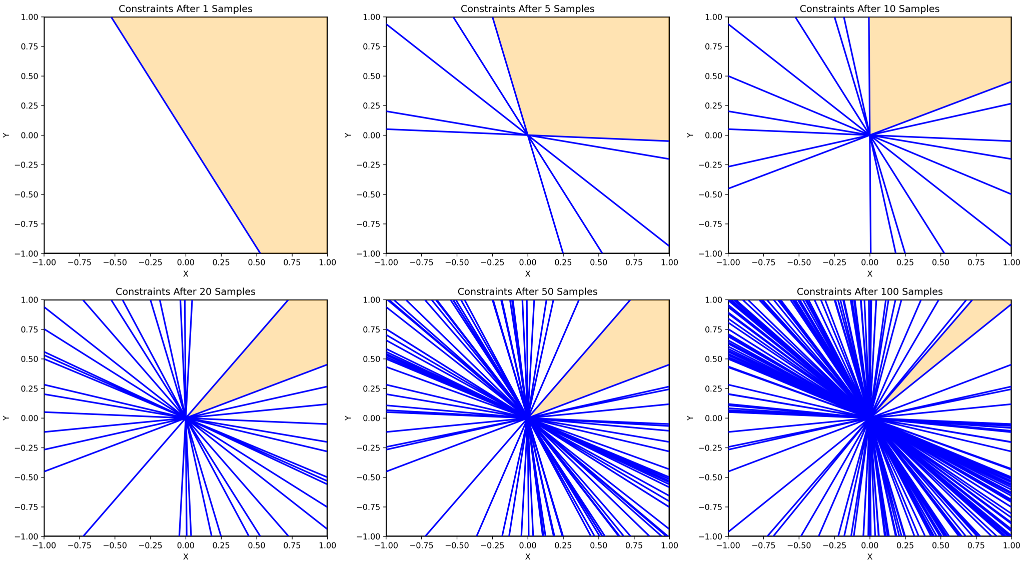
(a) Same Mean, Small Variance.



(b) Same Mean, Big Variance.



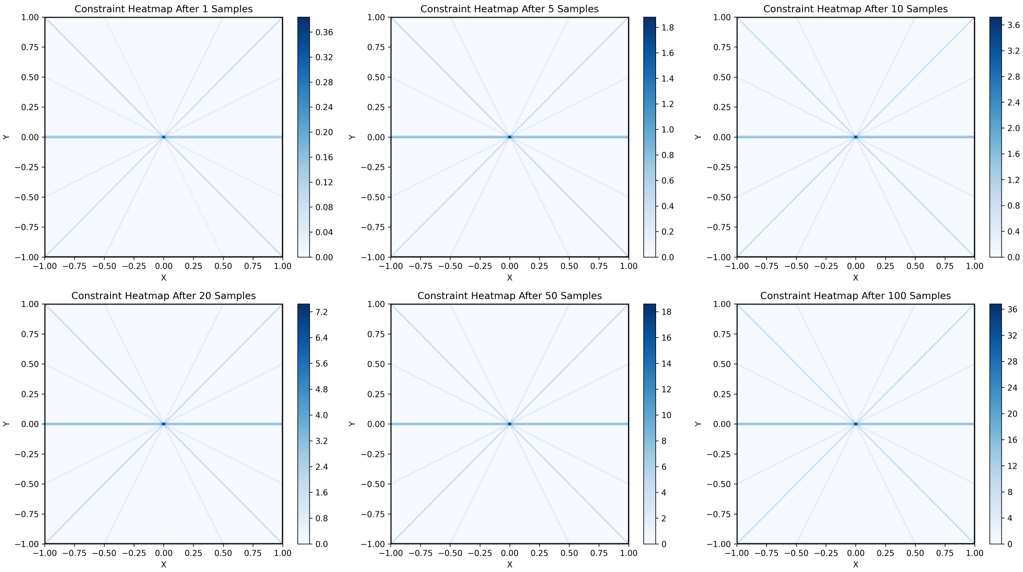
(c) Different Mean, Small Variance.



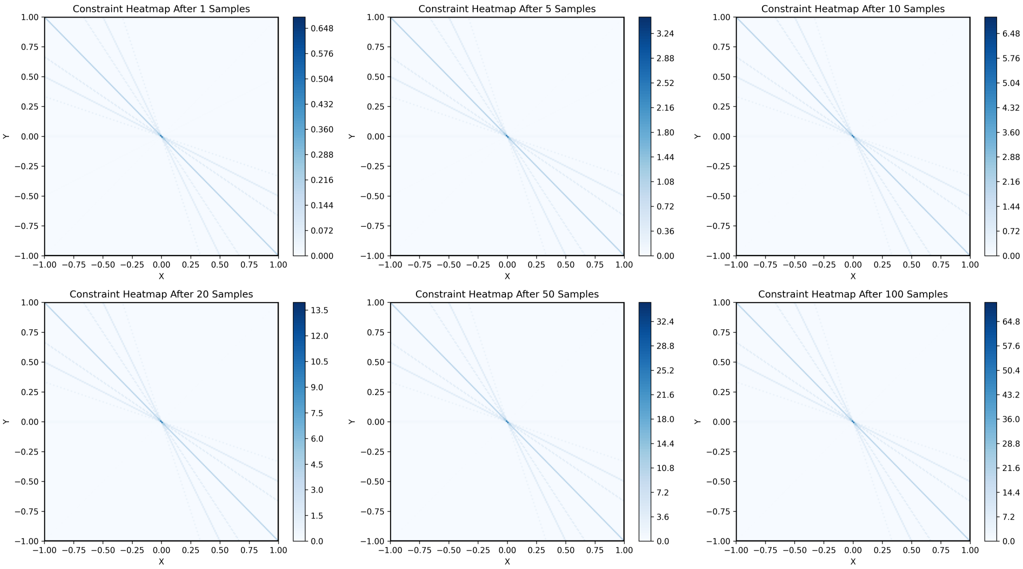
(d) Different Mean, Big Variance.

Figure 5.1: Visualization of the constraints in a run in the 4 configuration of the Gaussian.

Regarding the two binomial configurations, we considered that, given the constraints' limited and predefined probability of occurrence, an effective approach would be to conduct 1000 independent runs of our problem. Subsequently, we highlighted the most frequently observed constraints using a heatmap, where the intensity of the blue color in each line represents the frequency with which that constraint appeared across different runs. From Figure 5.2 we can see that only a predefined number of constraints are generated, some with higher probability than others.



(e) Same m.



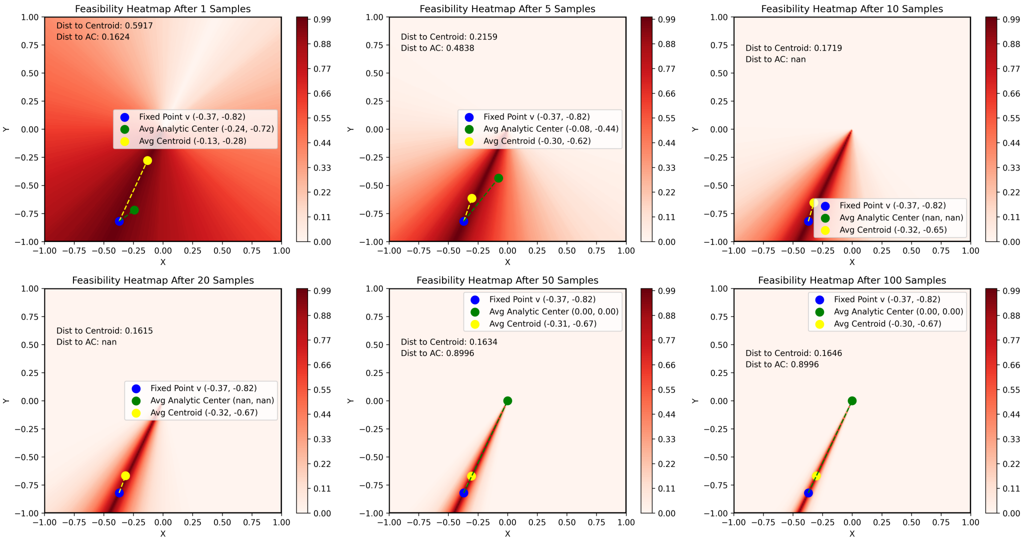
(f) Different m.

Figure 5.2: Visualization of the heatmap of the constraints in 1000 runs in the 2 configuration of the Binomial.

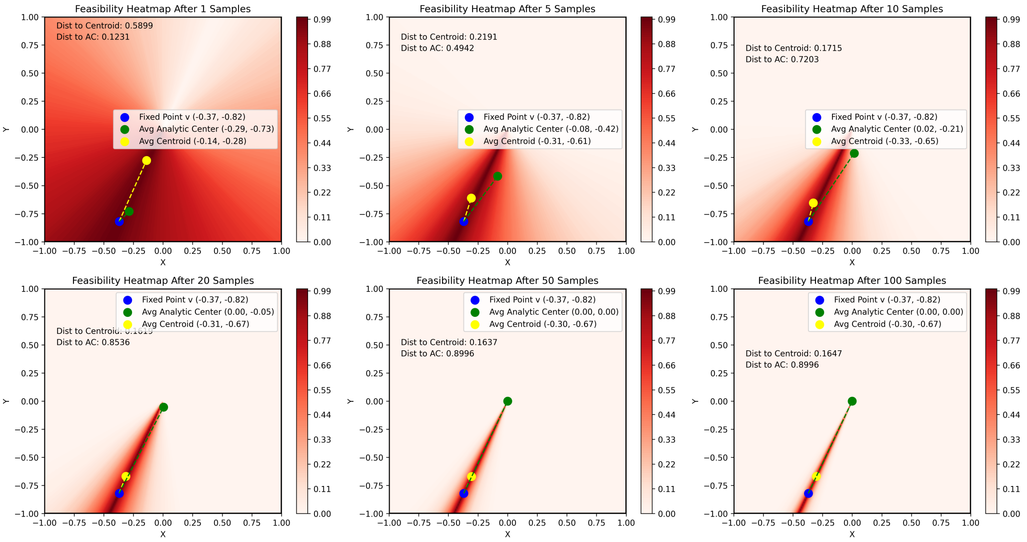
5.1.2. Shriking of the feasibility region

Following the visualization of the constraints, we proceeded to a second simulation focused on analyzing the contraction of the feasibility region $\mathbf{V}$. For this experiment, we simulated 1000 runs of our problem under each of the six configurations outlined in Section 5.1 and computed the average feasibility region for each sample size considered.

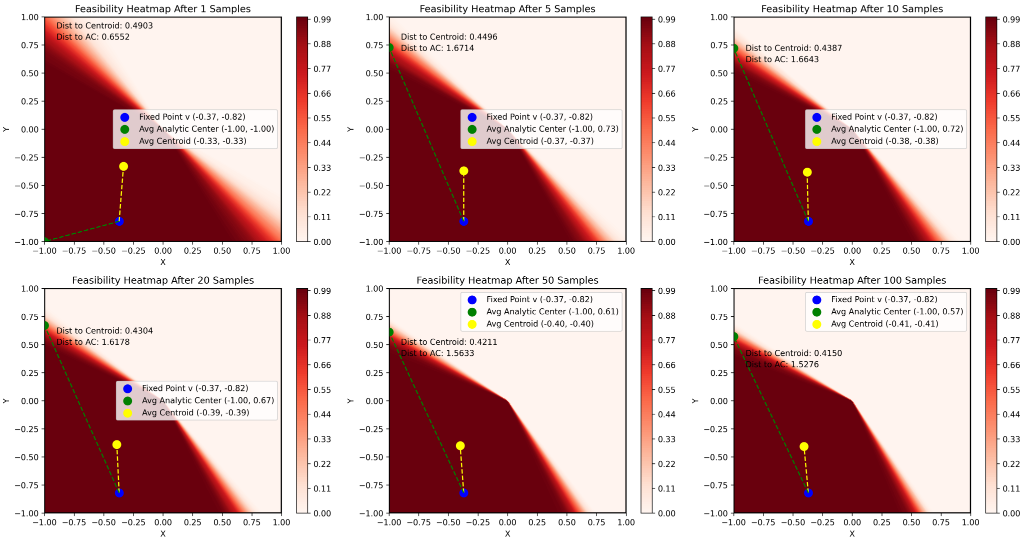
The subsequent figures illustrate the results. Figure 5.3 pertains to the two-dimensional case, where two candidate points were calculated as potential estimators for $\hat{\mathbf{v}}$, which approximates the optimal solution $\mathbf{v}^*$. These candidates include the Analytic Center [51], depicted in green, that is computed using Newton's method to find the point that is maximally distant from all constraint boundaries in the logarithmic sense, ensuring robustness even in ill-conditioned feasibility regions and the Mean Point, shown in yellow, which was ultimately selected as our $\hat{\mathbf{v}}$. Figure 5.4 extends the analysis to three-dimensional space to illustrate how the problem scales in higher dimensions. From their vision we can clearly understand how by expanding the dimensionality more samples are needed to shrink the $\mathbf{V}$ region and that in the case of an isotropic Gaussian it reduce faster.



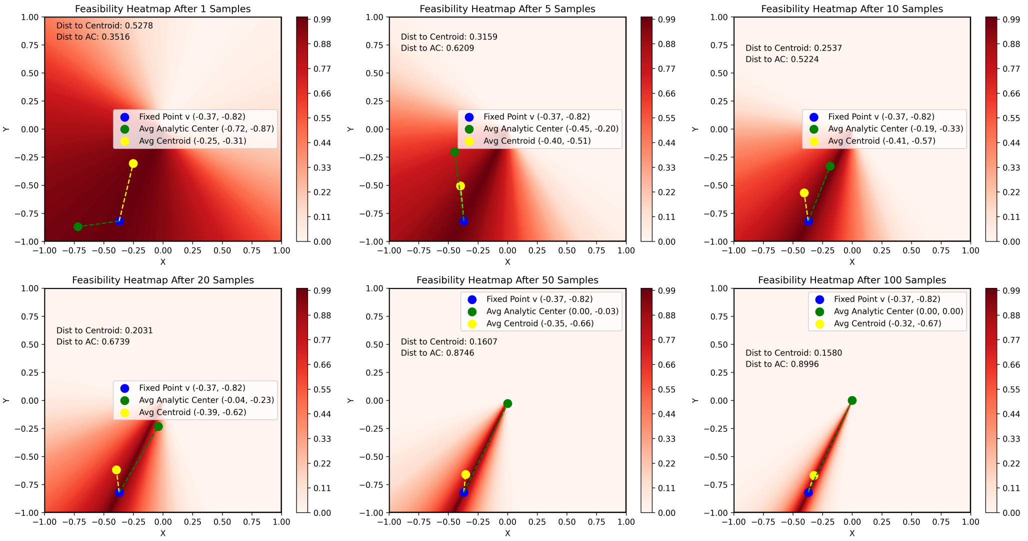
(a) Gaussian: Same Mean, Small Variance.



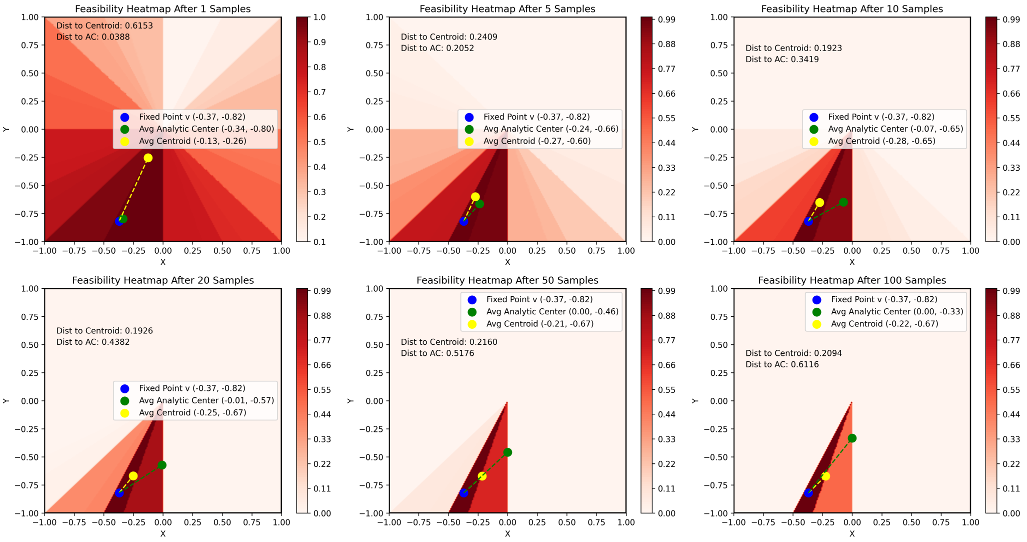
(b) Gaussian: Same Mean, Big Variance.



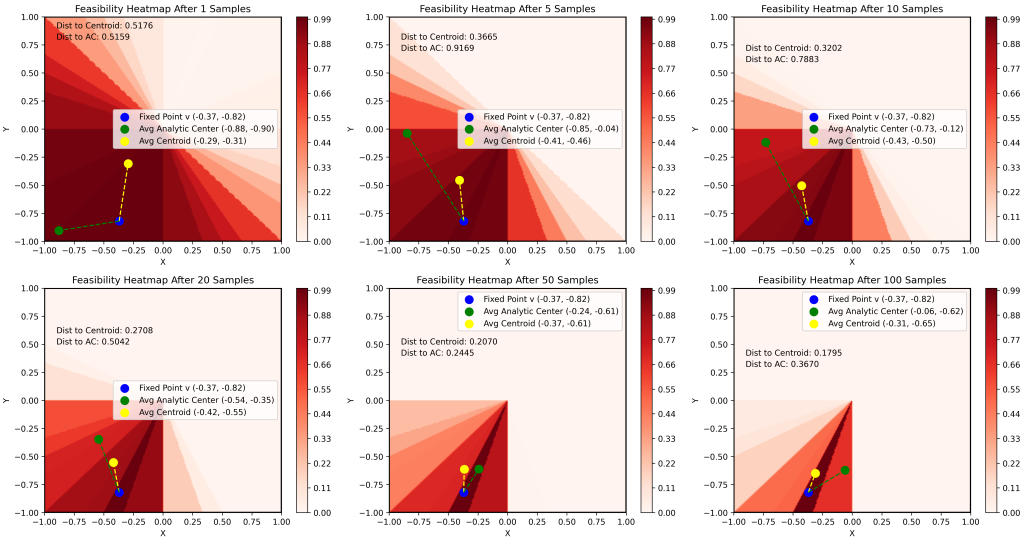
(c) Gaussian: Different Mean, Small Variance.



(d) Gaussian: Different Mean, Big Variance.

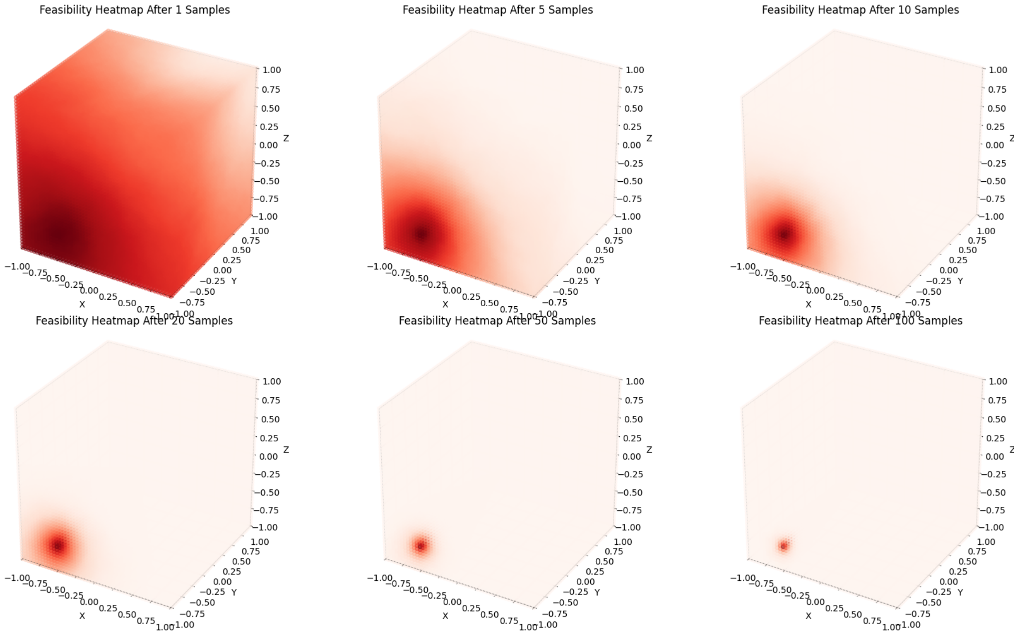


(e) Binomial: Same m.

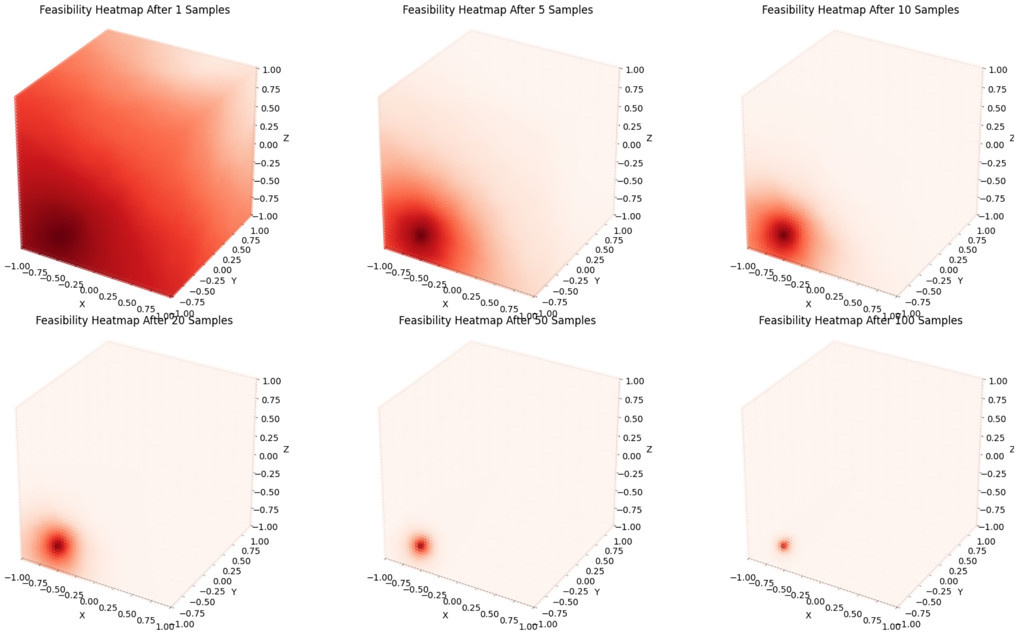


(f) Binomial: Different m.

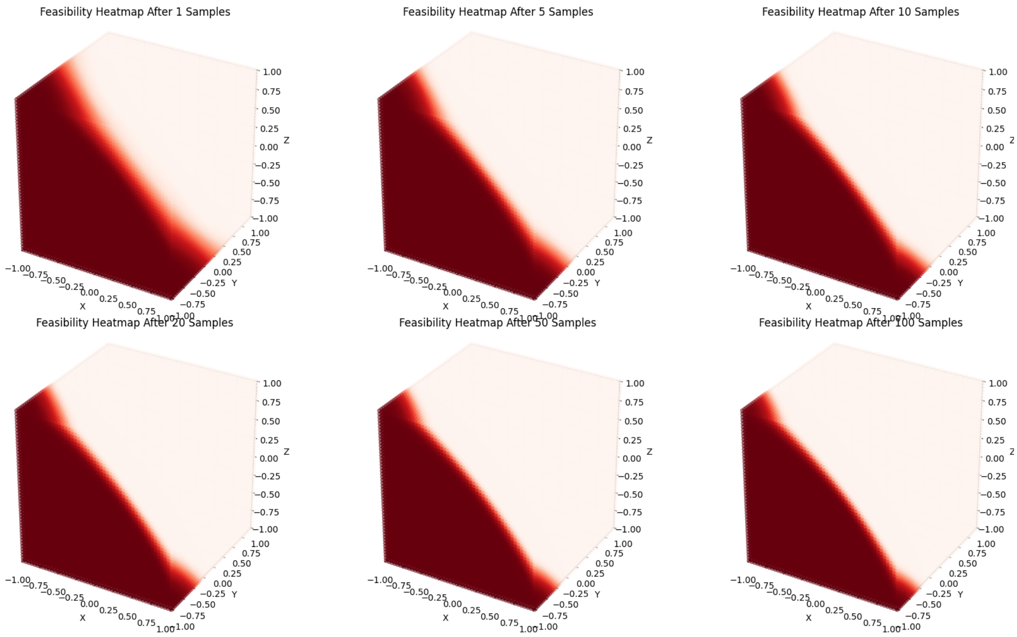
Figure 5.3: Visualization of the mean feasibility region for each sample size after 1000 runs in the 2d setting with the 6 configuration.



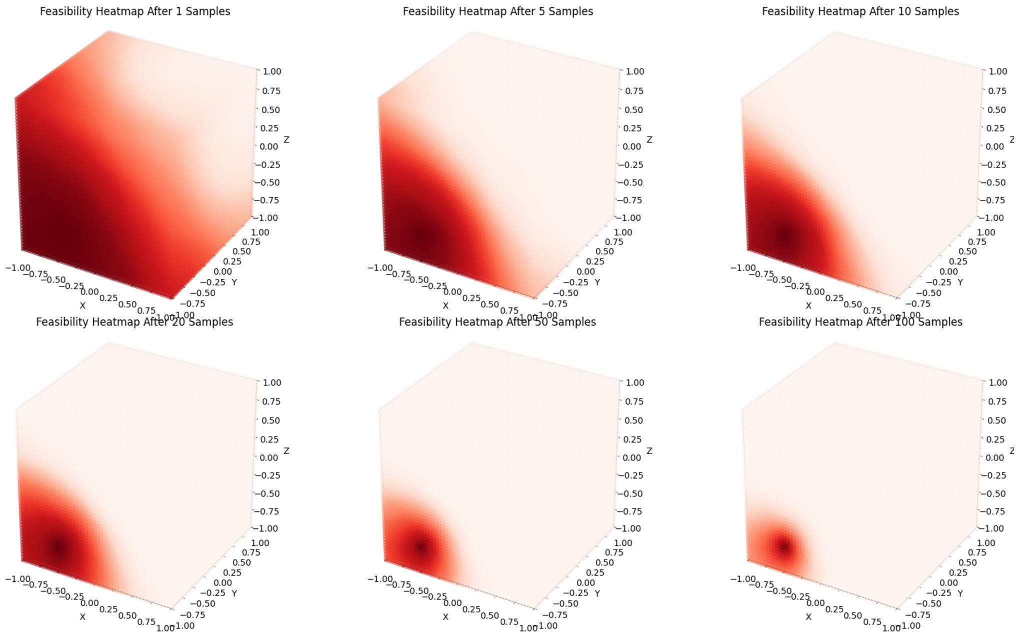
(a) Gaussian: Same Mean, Small Variance.



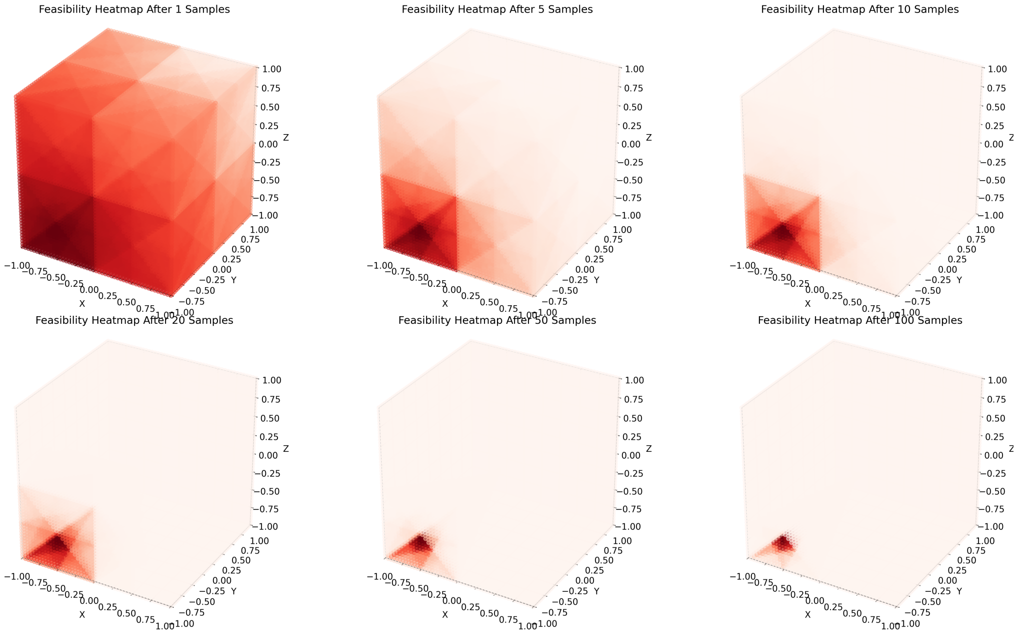
(b) Gaussian: Same Mean, Big Variance.



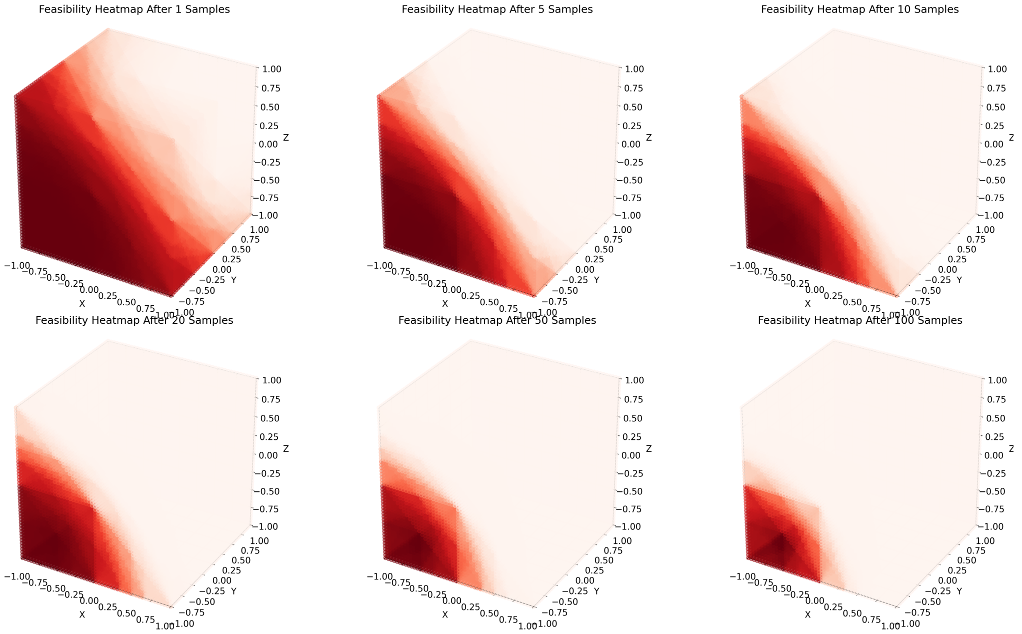
(c) Gaussian: Different Mean, Small Variance.



(d) Gaussian: Different Mean, Big Variance.



(e) Binomial: Same m.

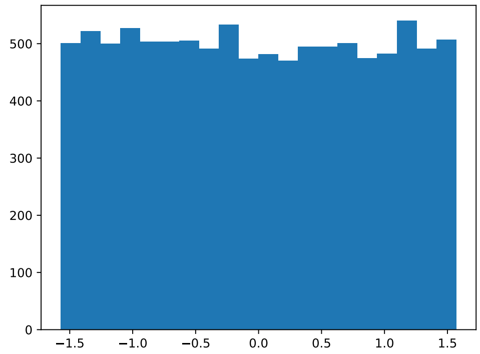


(f) Binomial: Different m.

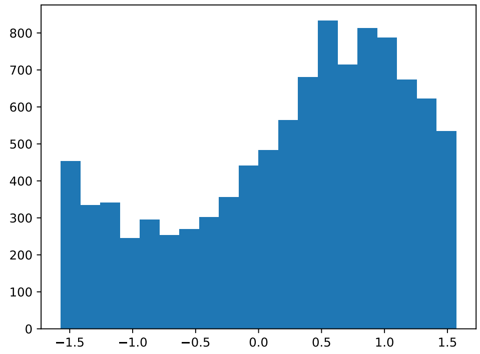
Figure 5.4: Visualization of the mean feasibility region for each sample size after 1000 runs in the 3d setting with the 6 configurations.

5.1.3. Arctangent of the Ratio of Different Distribution

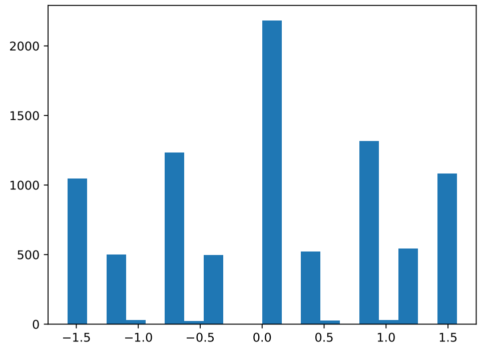
The final simulation in this section concerns the arctangent of the ratio of realizations from different distributions. In Lemma 3, we have theoretically demonstrated that for two isotropic Gaussian distributions, this ratio follows a uniform distribution. In this experiment, we extend our analysis by applying the same approach to additional configurations, including non-isotropic Gaussian and binomial distributions, to examine how deviations from isotropy and discreteness affect the observed behavior.



(a) Isotropic Gaussian.



(b) Non-Isotropic Gaussian.



(c) Binomial.

Figure 5.5: Distribution of the angle values computed as $\arctan(X/Y)$, where X and Y are sampled from the specified distribution. The y-axis represents the number of samples (out of 10,000) corresponding to each angle value.

5.1.4. Simulation Experiment Results

The simulations conducted in this section provided deeper insights into the problem. Notably, for a given sample size, feasibility regions in the three-dimensional case tend to be larger across all tested configurations. Moreover, it is evident that under Gaussian configurations, the feasibility region $\mathbf{V}$ contracts more rapidly until it reaches its minimal dimensionality (a hyperplane), particularly when $\mathbf{c}_1$ and $\mathbf{c}_2$ share the same mean, thereby eliminating bias in the sample generation process. Furthermore, the experimental results confirmed the uniformity of constraint generation in the case of the arctan of the ratio of two isotropic Gaussians. However, this uniformity does not persist in the non-isotropic (biased) cases or in the binomial case, where constraints are limited and discrete, as illustrated in Figure 5.5.

5.2. Preference Dissection Dataset Investigation

After establishing the theoretical framework of our problem through the application of simulation-based methodologies, we now transit to the core of the experimental analysis by conducting an in-depth examination of the dataset selected for evaluating the proposed algorithm.

The Preference Dissection dataset [50] comprises 5,240 observations distributed across 18 columns, each representing various AI-generated responses and their corresponding evaluations under different contextual scenarios. Below, we present a representative subset of the dataset to highlight its key characteristics.

Feature	**Type**	**Description**
query	Text	User input prompt
scenario_auto-j	Categorical	Task category (e.g., functional writing, code generation) classified by Auto-J's classifier
scenario_group	Categorical	Higher-level classification (e.g., Knowledge-aware)
response_1/response_2	Dictionary	AI-generated response and model that generate it, with the number of words determined by NLTK
gpt-4-turbo_reference	Text	Reference response from an advanced model

clear_intent	Boolean	Whether the user conveys intent clearly (Yes/No)
explicitly_express_feelings	Boolean	Whether the user conveys feelings explicitly (Yes/No)
explicit_constraints	List	Explicit constraints given in the prompt (e.g., "Be concise")
explicit_subjective_stances	List	Explicit subjective stance given in the query
explicit_mistakes_or_biases	List	Explicit mistakes given in the query
preference_labels	Dictionary	Preferred response for each judge (human or an LLM)
basic_response_1/ basic_response_2	Dictionary	Annotated rating by gpt-4 of the 20 basic properties for each response
errors_response_1/ errors_response_2	Dictionary	Categorization of response mistakes
query-specific_response_1/ query-specific_response_2	Dictionary	Annotation of the query properties

Table 5.1: Description of the dataset.

The primary objective of this dataset is to systematically assess and compare the performance of different AI-generated responses to user queries. This evaluation is conducted by categorizing the responses within predefined contexts and benchmarking them against reference responses. Each record in the dataset corresponds to a unique user query, which is classified according to its thematic and functional properties.

5.2.1. Preprocessing and Dataset Cleaning

The initial preprocessing step involves filtering out columns that are extraneous to our analysis. Only three fundamental columns serve as the basis for the experimental component of this study, thus justifying the selection of this dataset.

The first is `preference_labels`, which encapsulates the preferences expressed by 31 distinct LLMs, providing a comparative measure of response quality and a 32nd evaluation,

conducted by a human annotator.
The remaining two critical columns are `basic_response_1` and `basic_response_2`, which encode numerical values (ranging between 0 and 3) for 20 properties characterizing each response. From these two columns, we derived the feature vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ by converting the dictionary-structured properties into numerical arrays. From the `preference_labels` column, we extracted the human evaluation component, which serves as our target variable. Meanwhile, the AI-generated preference labels were transformed into a binary array of values $\{0, 1\}$.

Furthermore, we applied a filtering step to remove all instances in which the vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ were identical, ensuring that the dataset only contains cases where meaningful differences between responses exist. Finally, the dataset was partitioned into training and test subsets using an 80/20 split to facilitate model evaluation.

5.2.2. Feature Selection

Given the relatively limited size of the dataset and the dependency of sample complexity on the dimensionality of the feature space, we implemented a feature selection process [41] aimed at optimizing the representational efficiency of the input data. The selection process was guided by the correlation structure of the features, ensuring that the chosen features satisfy Assumption 4 regarding independence. Specifically, three correlation matrices were computed: one for the selected responses, one for the discarded responses, and one comparing the selected features with the discarded ones. Additionally, the feature selection process took into account the heterogeneity of feature values across different samples in the dataset.

Following this procedure, seven features were identified as the most informative for our analysis:

- **Authoritative Tone**: Measures the extent to which a response conveys confidence, expertise, and credibility in its delivery. Responses with a high authoritative tone typically use assertive language and structured argumentation to enhance perceived reliability.
- **Complex Word Usage and Sentence Structure**: Evaluates the lexical complexity and syntactic sophistication of a response. This includes factors such as the use of advanced vocabulary, varied sentence structures, and multi-clause constructions, which may contribute to readability and perceived intelligence.
- **Well-Formatted Responses**: Assesses the structural organization and visual clar-

ity of a response. Well-formatted responses include proper paragraphing, bullet points, headings, and other textual elements that enhance readability and comprehension.

- **Friendliness**: Captures the degree of warmth, approachability, and conversational engagement in a response. A friendly tone is characterized by polite phrasing, inclusive language, and an inviting conversational style.
- **Innovativeness and Novelty**: Measures the extent to which a response provides original, creative, or unconventional insights rather than standard or repetitive answers. This attribute is particularly relevant for assessing the generative capabilities of AI models in producing fresh perspectives.
- **Relevance (excluding accuracy considerations)**: Evaluates how well a response aligns with the user's query in terms of topicality and contextual appropriateness, irrespective of factual correctness. A highly relevant response addresses the core of the user's question while staying within the intended scope.
- **Information Richness (excluding inaccuracy considerations)**: Quantifies the depth and breadth of information provided in a response, independent of its factual accuracy. A response with high information richness includes comprehensive explanations, supporting details, and nuanced insights that contribute to its informativeness.

Appendix A presents the correlation matrices for the selected features, as well as histograms of the vector $\mathbf{c}_{\text{selected}} - \mathbf{c}_{\text{discarded}}$, where the first vector contains the contexts of the answer preferred by the user meanwhile the other the ones discarded. A Gaussian distribution has been fitted to approximate the underlying data-generating distribution. Appendix A also provides the correlation matrix for all 20 features, along with histograms of the vectors $\mathbf{c}_{\text{selected}}$ and $\mathbf{c}_{\text{discarded}}$, with the best-fitting Gaussian and binomial distributions superimposed.

5.3. Finding a Selected $\mathbf{v}^*$ in the Dataset

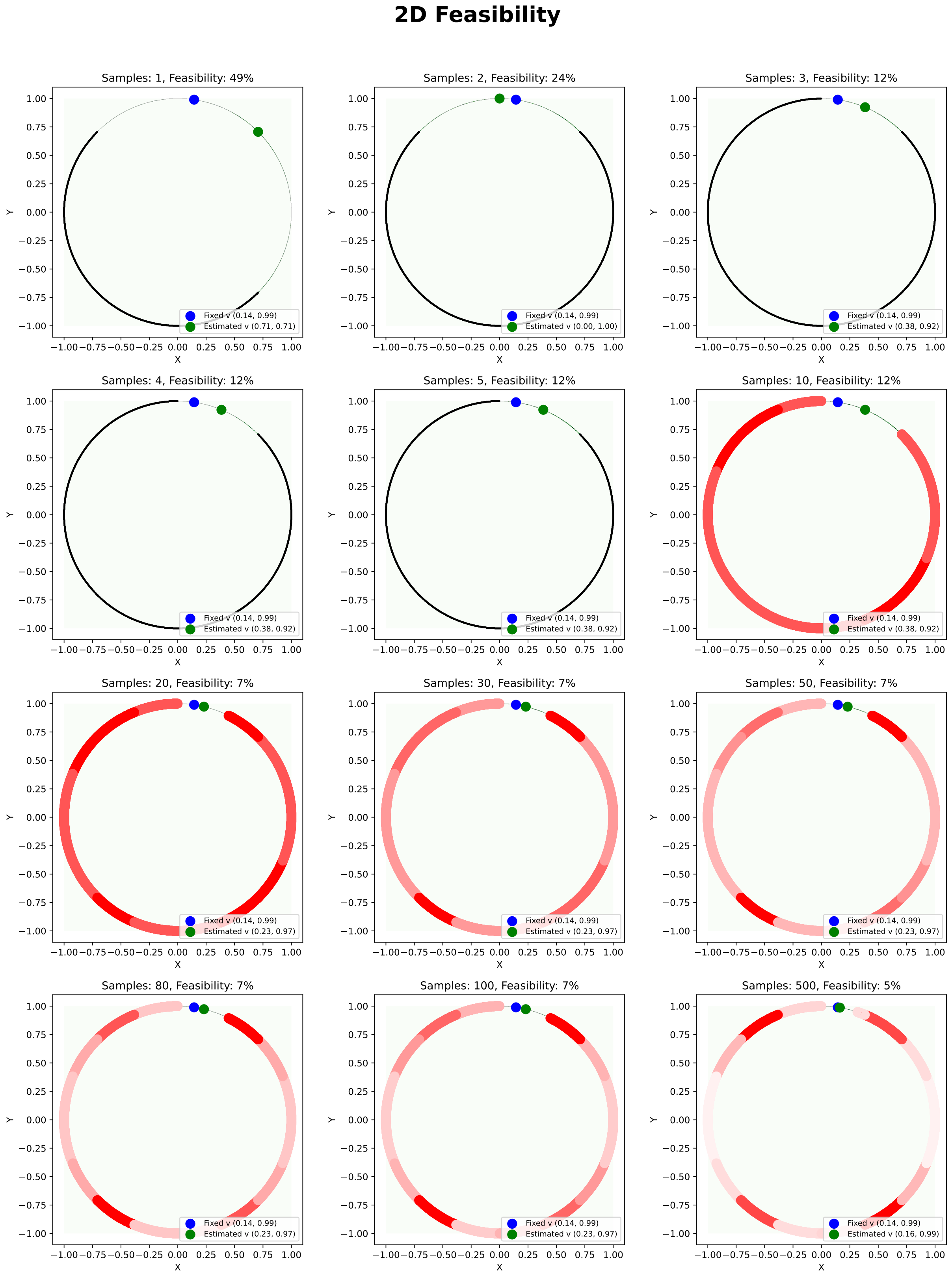
In this section, we conducted a series of experiments using hybrid data, consisting of both real samples from the dataset and artificially generated samples derived from fitted distributions. Specifically, we generated new data points based on the statistical properties of selected categories, yielding four new vectors: $\mathbf{c}_{1,g}$ and $\mathbf{c}_{2,g}$ (obtained from Gaussian fits), as well as $\mathbf{c}_{1,b}$ and $\mathbf{c}_{2,b}$ (obtained from binomial fits).

The objective of this experiment is to analyze the impact of using different representations of the response vectors in reconstructing preferences. To this end, we compared the performance of the system when using different variations of $\mathbf{c}$ while keeping a fixed reference vector $\mathbf{v}^*$, which is manually selected. The preference is then reconstructed on a sample-by-sample basis according to the corresponding values of $\mathbf{c}_1$ and $\mathbf{c}_2$.

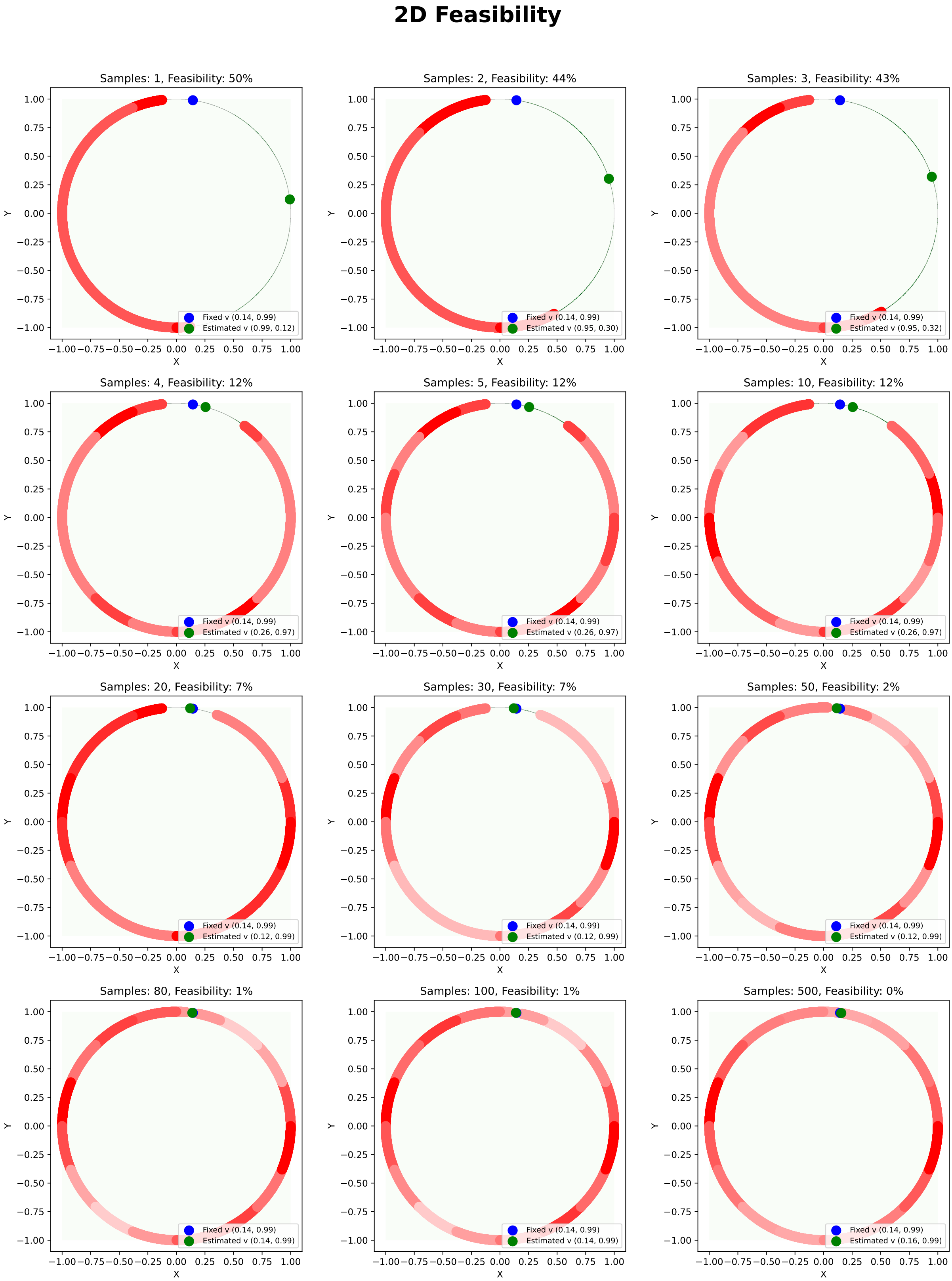
In the first set of visualizations, we analyzed the two-dimensional case, where the selected response characteristics are *Innovativeness and Novelty* and *Relevance (excluding accuracy considerations)*. Figure 5.6 illustrates how the feasibility region (highlighted in green) is structured and how constraints are distributed across different sectors of a circular representation (divided into eight equal parts). The intensity of the red coloration in each sector indicates the density of samples within that region—darker shades correspond to higher sample concentrations.

To ensure statistical robustness, we repeated this experiment 100 times, permuting the dataset in each run. This was necessary because the sequential order in which samples are processed influences the partitioning of the feasibility region. The final results were obtained by averaging across all runs.

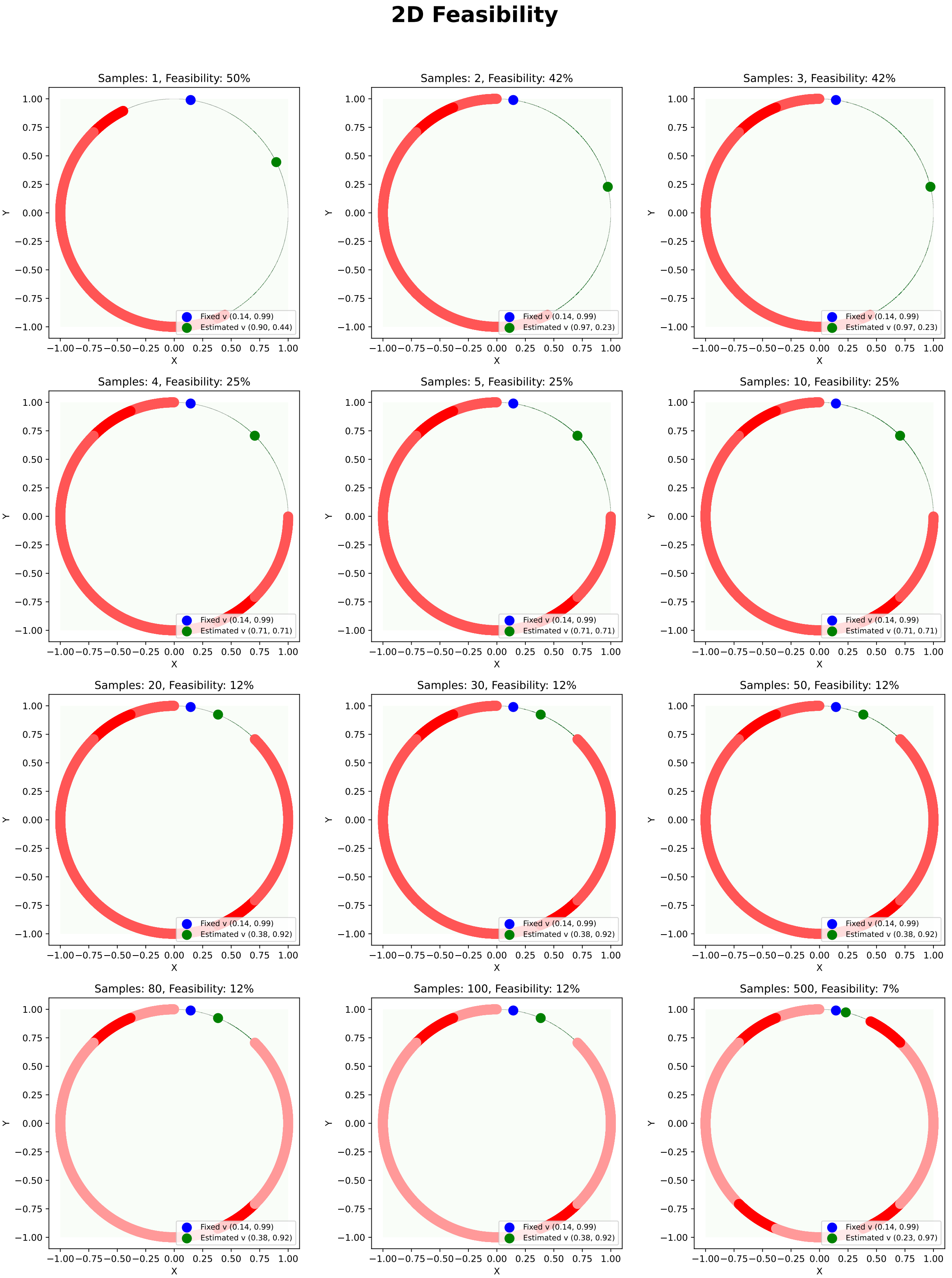
Additionally, two supplementary graphs provide further insights. Figure 5.7 illustrates the percentage change in the feasibility area as new samples are introduced, while Figure 5.8 quantifies the variation in the distance between the fixed reference vector $\mathbf{v}$ and the estimated preference vector $\hat{\mathbf{v}}$. These two metrics are directly related to the sample complexity considerations discussed in the previous chapter. The x-axis, which represents the number of evaluated samples, is displayed on a logarithmic scale. This choice is motivated by the observation that as the number of processed samples increases, further reductions in the feasibility area become increasingly difficult to achieve. Consequently, the earliest samples contribute more significantly to shaping the feasibility region. The same experimental procedure is extended to the three-dimensional case by incorporating a third response characteristic, *Complex Word Usage and Sentence Structure*. The results of these 3d experiments are presented in Figures 5.9, 5.10, and 5.11.


2D Feasibility
Samples: 1, Feasibility: 49%
Fixed v (0.14, 0.99)
Estimated v (0.71, 0.71)
Samples: 2, Feasibility: 24%
Fixed v (0.14, 0.99)
Estimated v (0.00, 1.00)
Samples: 3, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.38, 0.92)
Samples: 4, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.38, 0.92)
Samples: 5, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.38, 0.92)
Samples: 10, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.38, 0.92)
Samples: 20, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.23, 0.97)
Samples: 30, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.23, 0.97)
Samples: 50, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.23, 0.97)
Samples: 80, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.23, 0.97)
Samples: 100, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.23, 0.97)
Samples: 500, Feasibility: 5%
Fixed v (0.14, 0.99)
Estimated v (0.16, 0.99)
X
Y
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

(a) Binomial.

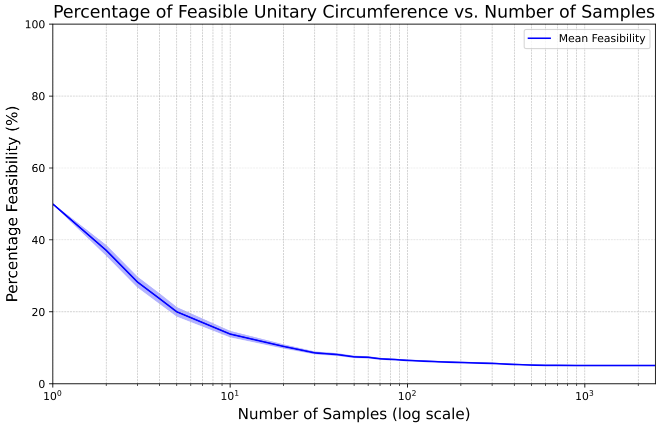

2D Feasibility
Samples: 1, Feasibility: 50%
Fixed v (0.14, 0.99)
Estimated v (0.99, 0.12)
Samples: 2, Feasibility: 44%
Fixed v (0.14, 0.99)
Estimated v (0.95, 0.30)
Samples: 3, Feasibility: 43%
Fixed v (0.14, 0.99)
Estimated v (0.95, 0.32)
Samples: 4, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.26, 0.97)
Samples: 5, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.26, 0.97)
Samples: 10, Feasibility: 12%
Fixed v (0.14, 0.99)
Estimated v (0.26, 0.97)
Samples: 20, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.12, 0.99)
Samples: 30, Feasibility: 7%
Fixed v (0.14, 0.99)
Estimated v (0.12, 0.99)
Samples: 50, Feasibility: 2%
Fixed v (0.14, 0.99)
Estimated v (0.12, 0.99)
Samples: 80, Feasibility: 1%
Fixed v (0.14, 0.99)
Estimated v (0.14, 0.99)
Samples: 100, Feasibility: 1%
Fixed v (0.14, 0.99)
Estimated v (0.14, 0.99)
Samples: 500, Feasibility: 0%
Fixed v (0.14, 0.99)
Estimated v (0.16, 0.99)
X
Y
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

(b) Gaussian.

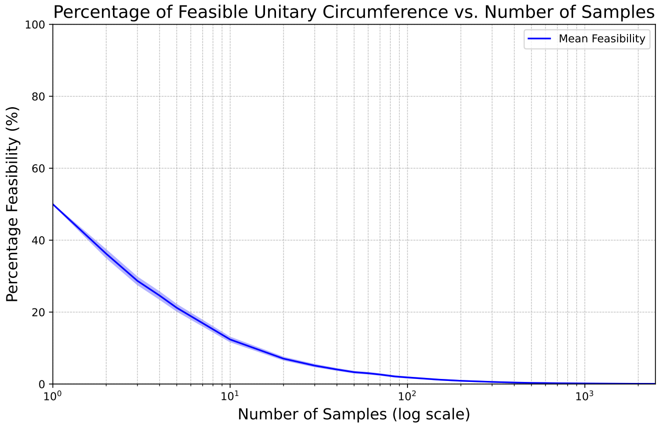


(c) Original Dataset.

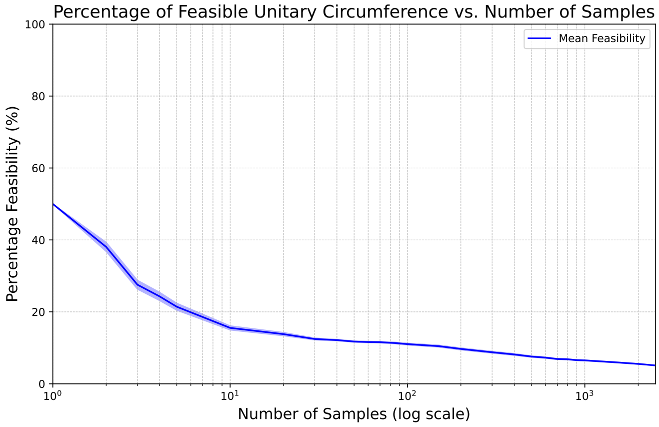
Figure 5.6: Visualization of the feasibility area in 2d of different distribution.



(a) Binomial.

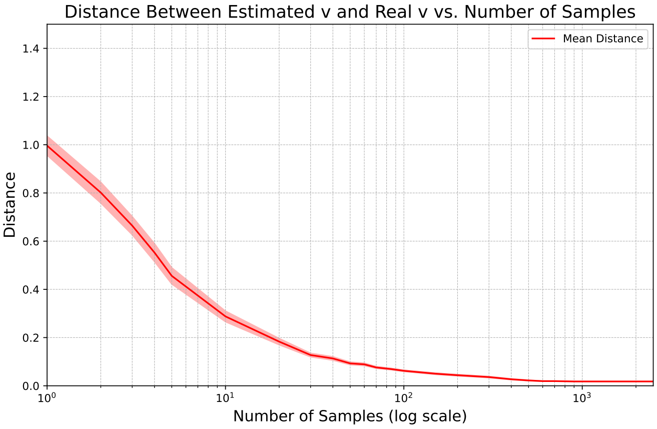


(b) Gaussian.

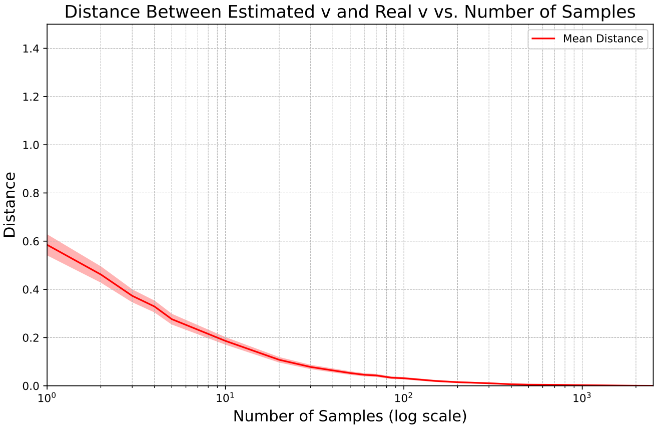


(c) Original Dataset.

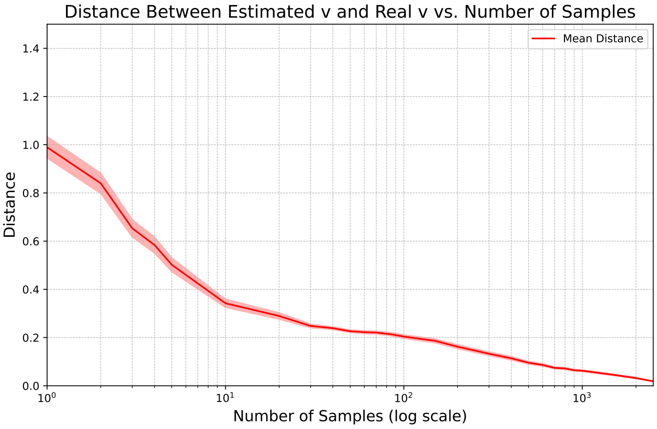
Figure 5.7: Percentage of the circle in the feasibility area in 2d vs number of samples with 100 permutations of the dataset. The mean distance is shown in bold, with the standard deviation highlighted in a lighter color.



(a) Binomial.

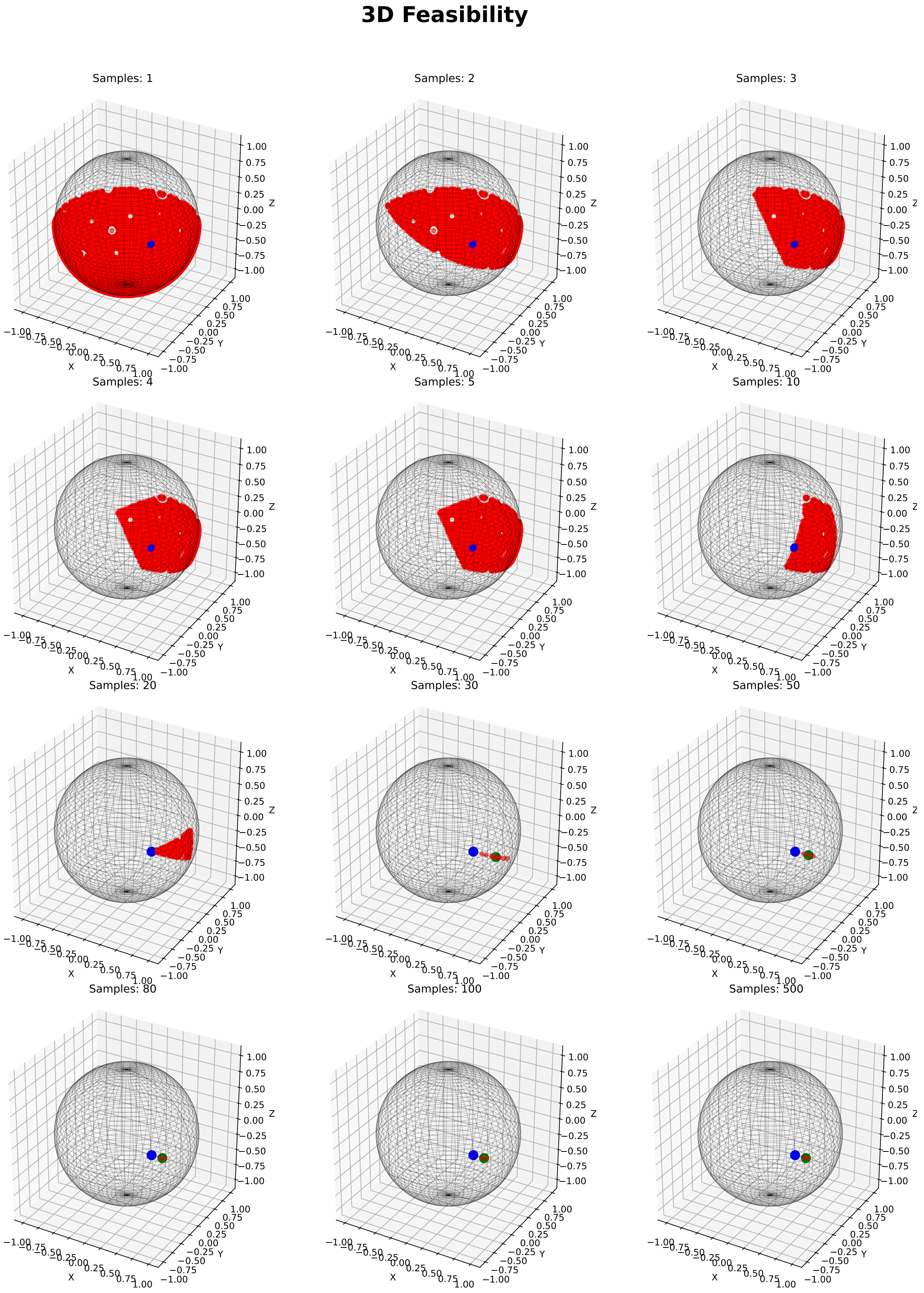


(b) Gaussian.

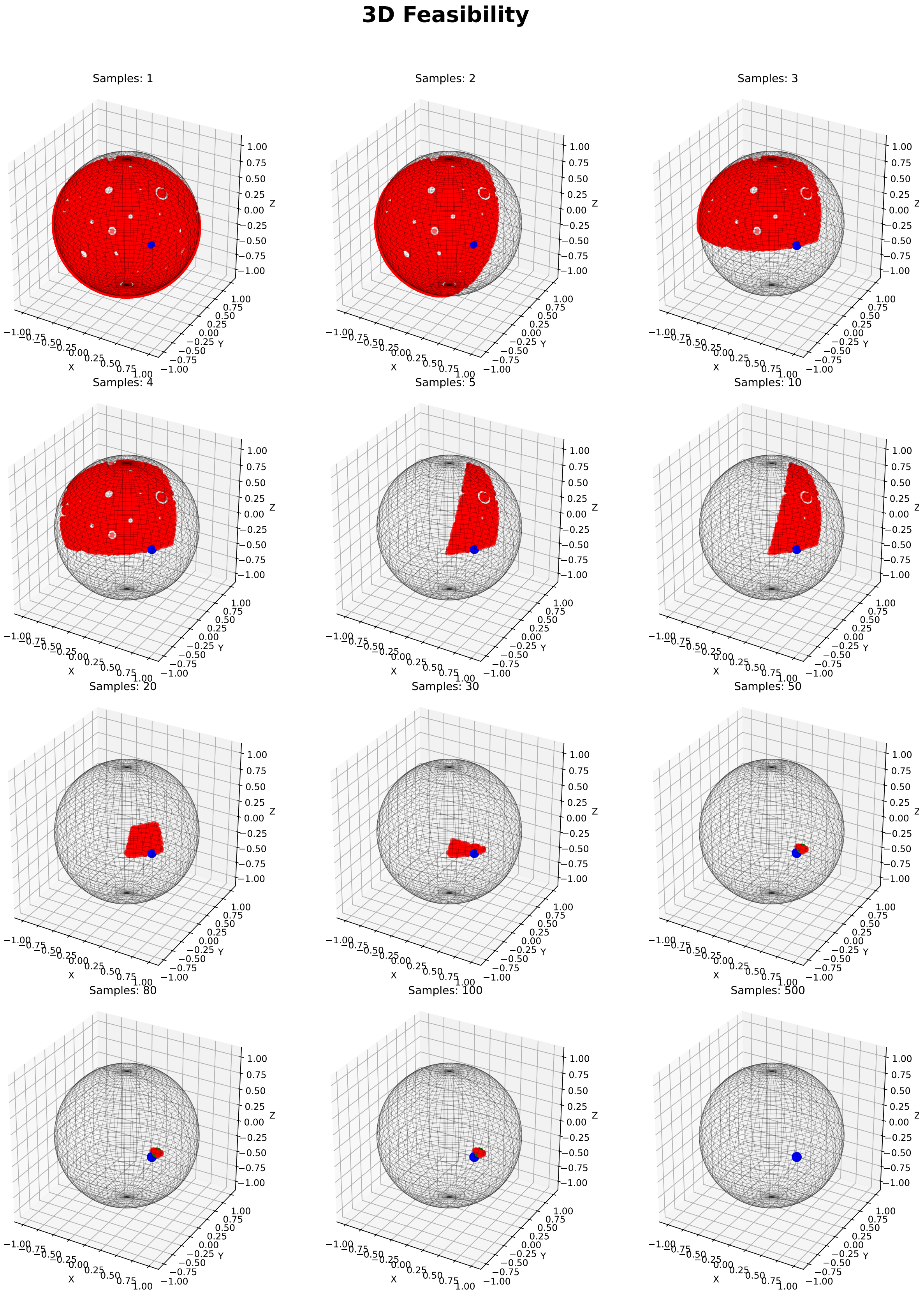


(c) Original Dataset.

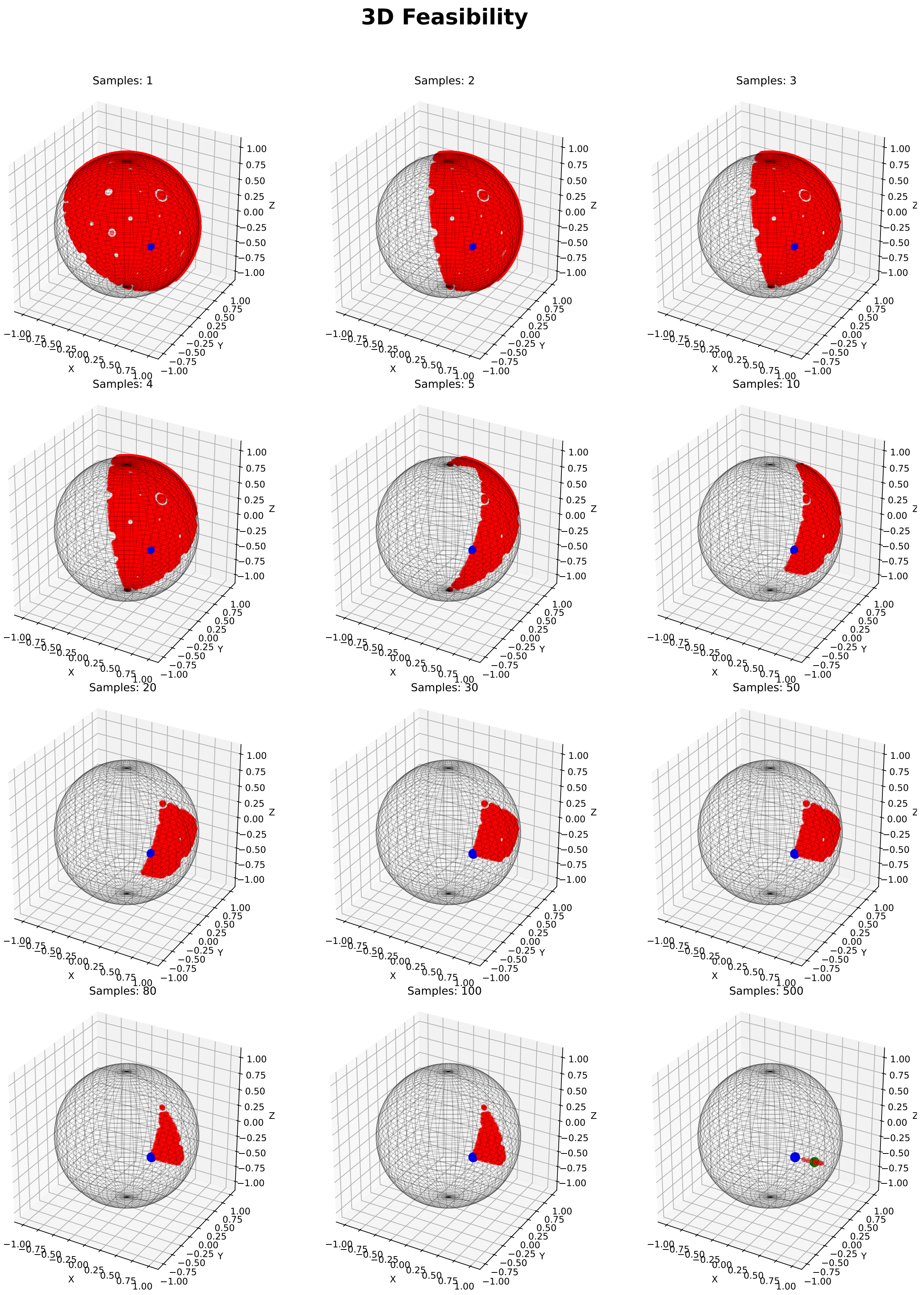
Figure 5.8: Distance between $\hat{\mathbf{v}}$ and $\mathbf{v}^*$ in 2d vs number of samples with 100 permutations of the dataset. The mean distance is shown in bold, with the standard deviation highlighted in a lighter color.


3D Feasibility
Samples: 1
Samples: 2
Samples: 3
Samples: 4
Samples: 5
Samples: 10
Samples: 20
Samples: 30
Samples: 50
Samples: 80
Samples: 100
Samples: 500
X
Y
Z
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

(a) Binomial.

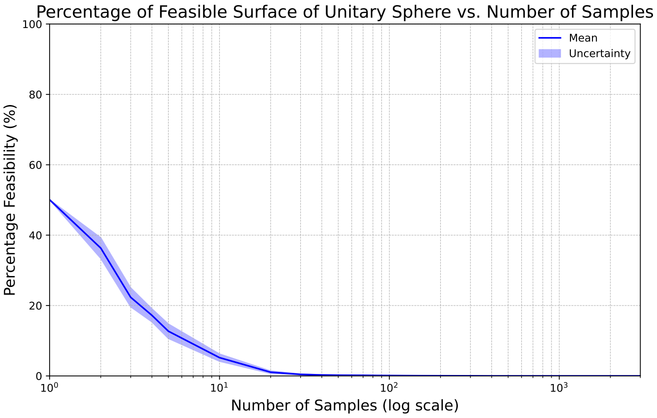

3D Feasibility
Samples: 1
Samples: 2
Samples: 3
Samples: 4
Samples: 5
Samples: 10
Samples: 20
Samples: 30
Samples: 50
Samples: 80
Samples: 100
Samples: 500
X
Y
Z
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

(b) Gaussian.

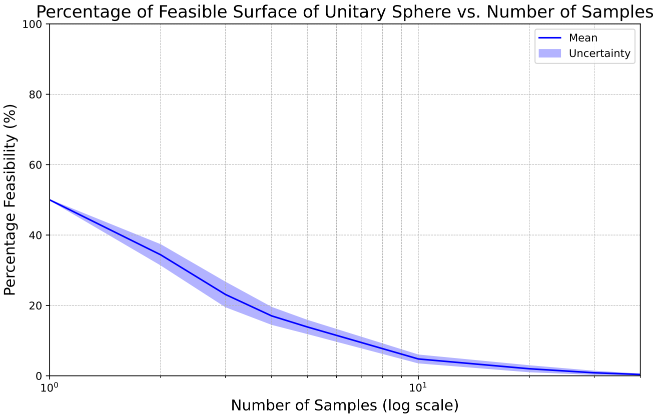


(c) Original Dataset.

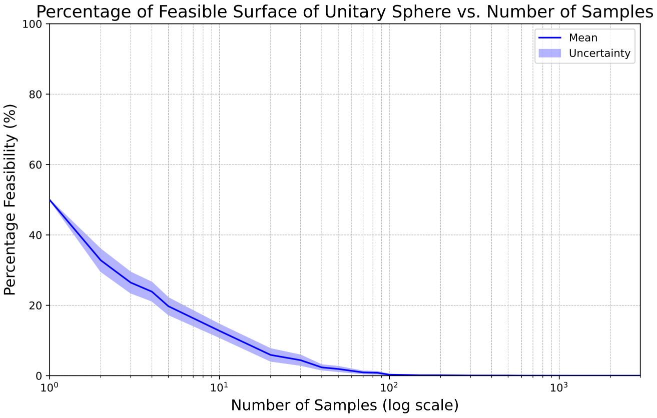
Figure 5.9: Visualization of the feasibility area in 3d of different distribution.



(a) Binomial.

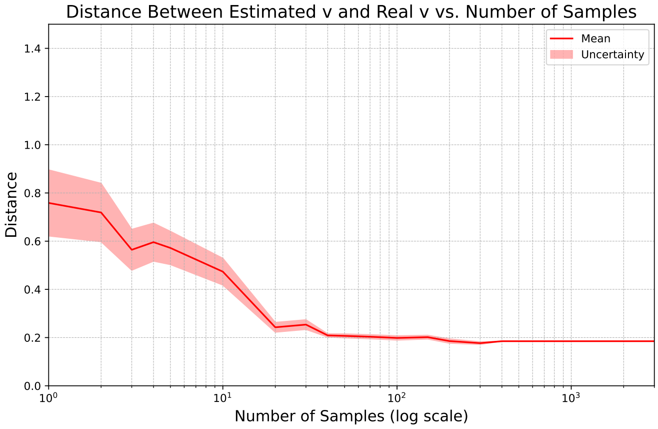


(b) Gaussian.

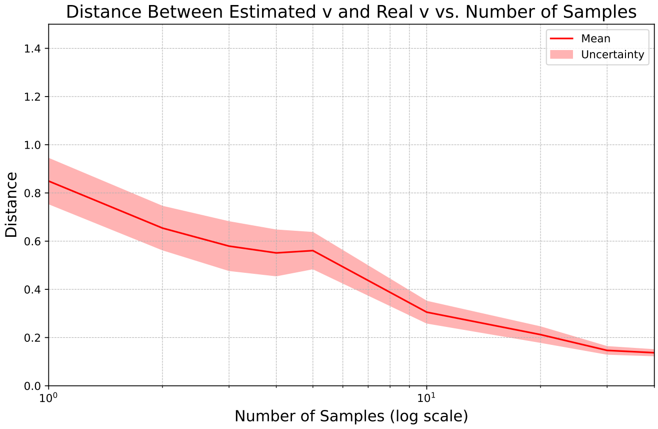


(c) Original Dataset.

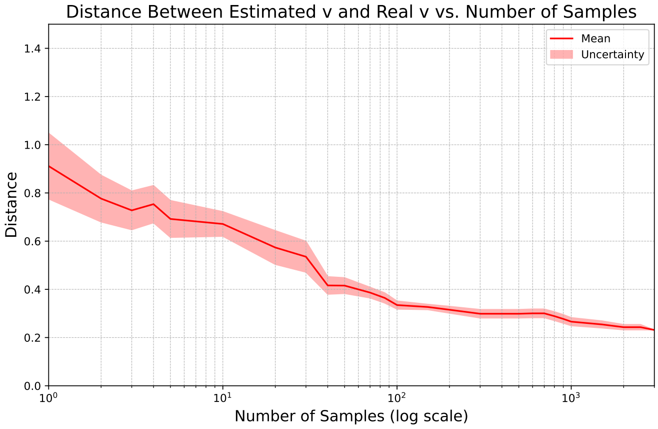
Figure 5.10: Percentage of the circle in the feasibility area in 3d vs number of samples with 10 permutations of the dataset. The mean distance is shown in bold, with the standard deviation highlighted in a lighter color.



(a) Binomial.



(b) Gaussian.



(c) Original Dataset.

Figure 5.11: Distance between $\hat{\mathbf{v}}$ and $\mathbf{v}^*$ in 3d vs number of samples with 10 permutations of the dataset. The mean distance is shown in bold, with the standard deviation highlighted in a lighter color.

5.3.1. Result of the Comparison of Various Distribution

As expected, in the two-dimensional case, we achieve a smaller average distance compared to its three-dimensional counterpart. Another key observation from the visual representations is that the only distribution capable of precisely identifying the fixed point is the Gaussian distribution. This outcome aligns well with the theoretical framework and the inherent properties of the distribution. It is notable to highlight that in all three experimental settings, the initial feasibility region is effectively constrained using only a small subset of the dataset.

Furthermore, the Gaussian distribution exhibits the highest standard deviation. This phenomenon arises because, given the uniformity of the constraints in the feature space, the order in which samples are processed plays a crucial role in determining the final distribution.

A less intuitive but notable result is that our experimental setting is also well-aligned with the structural characteristics of the original dataset. Specifically, the results obtained using the binomial distribution closely resemble those derived from the original dataset, indicating that our generative approach successfully captures key contextual properties of the data.

5.4. Test the Cutting Plane Preference Learner Algorithm

As a final experiment, we assess the predictive performance of the Cutting Plane Preference Learner algorithm. To achieve this, we employed the following experimental setup. The circumference was divided into eight equal segments, from which we derived eight fixed points $\mathbf{v}^*$, corresponding to eight distinct phases.

For each phase, we reconstructed the preference vector associated with the respective fixed point and trained our algorithm using the training portion of the dataset. The trained model was then evaluated on the test set to assess its generalization capabilities. As a baseline, we employed a majority voting algorithm, which aggregated the preference vectors of the 31 LLMs extracted from the original dataset.

The results of this evaluation are summarized in Table 5.2 and Table 5.3.

v^* (coordinates)	Majority Voting		
	Accuracy	Precision	Recall
$(1; 0.01)$	0.59	0.62	0.65
$(0.71; 0.71)$	0.56	0.63	0.62
$(0.01; 1)$	0.50	0.53	0.58
$(-0.7; 0.72)$	0.25	0.26	0.32
$(-1; 0.01)$	0.31	0.30	0.38
$(-0.71; -0.71)$	0.47	0.49	0.55
$(0.01; -1)$	0.75	0.74	0.81
$(0.72; -0.70)$	0.69	0.70	0.75

Table 5.2: Majority Voting Performance.

v^* (coordinates)	Cutting Plane Preference Learner				
	Accuracy	Precision	Recall	$\hat{v}$ (coordinates)	Distance
$(1; 0.01)$	1	1	1	$(0.92; 0.38)$	0.38
$(0.71; 0.71)$	1	1	1	$(0.71; 0.71)$	0
$(0.01; 1)$	1	1	1	$(0.16; 0.99)$	0.15
$(-0.7; 0.72)$	1	1	1	$(-0.59; 0.81)$	0.15
$(-1; 0.01)$	1	1	1	$(-0.92; 0.38)$	0.38
$(-0.71; -0.71)$	1	1	1	$(-0.71; -0.71)$	0
$(0.01; -1)$	1	1	1	$(0.2; -0.97)$	0.22
$(0.72; -0.70)$	1	1	1	$(0.92; -0.38)$	0.37

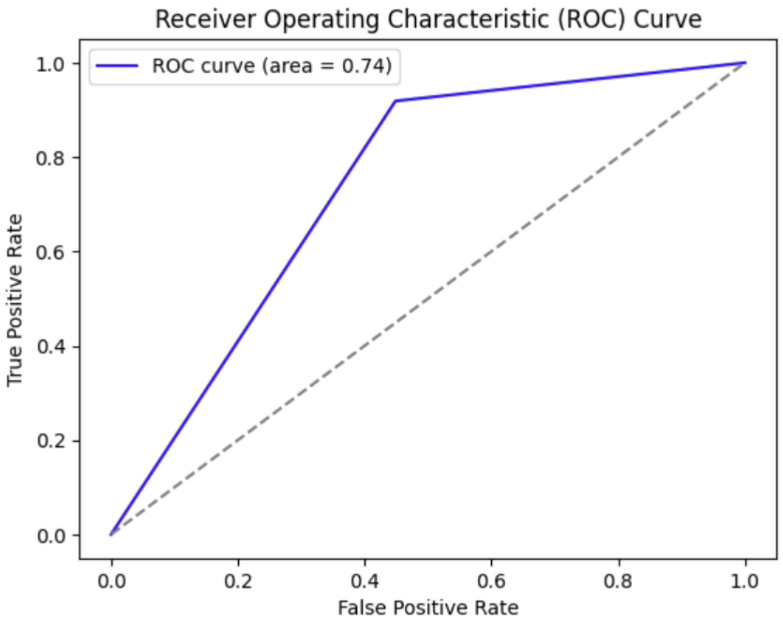
Table 5.3: Cutting Plane Preference Learner Performance.

5.4.1. Result of the Cutting Plane Preference Learner algorithm

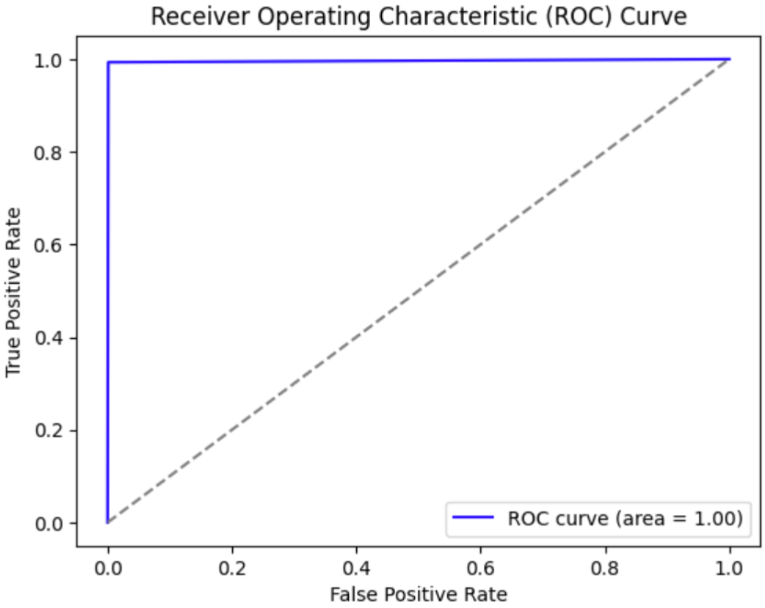
As evidenced by Table 5.2 and Table 5.3, our algorithm demonstrates desirable approximation capabilities in the two-dimensional case, effectively estimating the fixed point $\mathbf{v}^*$ within each sector. Notably, even in sectors where the distance between the estimated and actual points increases, the model still accurately captures the underlying structural patterns of the data.

In contrast, the majority voting approach, despite leveraging the combined expressiveness of 31 LLMs, exhibits systematic biases with respect to certain regions of the circumference. This suggests that while ensemble methods can enhance robustness, they may also introduce inherent preferences that affect performance in specific cases.

To conclude our analysis, we conducted an additional experiment incorporating all seven selected response categories. The resulting Receiver Operating Characteristic (ROC) curves are presented below in Figure 5.12. While the Cutting Plane Preference Learner Algorithm achieves a near-perfect performance with an Area Under the Curve (AUC) of 100%, in accordance with the statement in Assumption 1, we consider a setting in which the space is linearly separable. Consequently, we anticipated that the application of the algorithm would effectively identify the region containing the fixed point $\mathbf{v}^*$. For the first time, we observe instances of misclassification in the confusion matrix as we can seen in Table 5.7. This observation suggests that additional training samples may be necessary to extend the vector representation of contexts, thereby enhancing the model's generalization capacity.



(a) Majority Voting.



(b) Cutting Plane Preference Learner.

Figure 5.12: ROC curve in 7d experiment.

Metric	Value
Accuracy	0.6033
Precision	0.2523
Recall	0.9189
ROC AUC	0.7351

Table 5.4: Evaluation Metrics Majority Voting.

		Predicted	
		p0	**p1**
Actual	**p0**	495	403
	p1	12	136

Table 5.5: Confusion Matrix Majority Voting.

Metric	**Value**
Accuracy	0.9981
Precision	0.9932
Recall	0.9932
ROC AUC	0.9961

Table 5.6: Evaluation Metrics Cutting Plane Preference Learner.

		Predicted	
		p0	**p1**
Actual	**p0**	897	1
	p1	1	147

Table 5.7: Confusion Matrix Cutting Plane Preference Learner.

6 | Conclusions

In this thesis, we investigated the problem of selecting between two alternatives within the context of response alignment for LLMs. This problem was formulated as an offline feasibility problem, wherein the objective was to delineate the user's preference space by leveraging the context vectors associated with the two options as constraints.

Furthermore, we modeled these constraints using three distinct distributions, namely isotropic Gaussian, non-isotropic Gaussian, and binomial distributions. For each instance of the problem, we provided theoretical guarantees regarding its resolution through the formulation of sample complexity. To reinforce these theoretical findings, we conducted an extensive simulation-based experimental campaign, which empirically substantiated our theoretical results.

Additionally, we introduced an algorithm, *The Cutting Plane Preference Learner*, which was designed to manage the entire pipeline for aligning an LLM with human preferences. We developed a series of validation tests to ensure its correct functioning. These tests were executed on a real dataset, which we modified with synthetic data that adhered to our initial assumptions. The outcomes of our experiments were then compared against a majority voting approach, wherein the judgment of 31 distinct LLMs was aggregated. The results, which aligned with our theoretical guarantees, demonstrated that in our specific setting, the proposed algorithm outperformed the majority voting approach. In particular, it effectively identified the region of the space in which the human preference vector resided, even in high-dimensional scenarios.

A limitation of our approach stemmed from the constrained number of samples available for validation, which restricted our ability to extend the analysis to higher spaces. Nevertheless, in comparison to the existing state of the art, the research direction pursued in this work appeared promising, particularly concerning its sample efficiency in learning human preferences throughout the entire space, which added explainability to the whole process.

6.1. Future Works

In conclusion, we present several possible directions to extend our work. Our algorithm assumes linearity between the context vectors and the fixed $\mathbf{v}^*$, a first future research could focus on adapting it to non-linear scenarios that more accurately reflect the complexity of existing applications.

Another significant extension regards the modeling of user preference. It is possible to improve the realism of the scenario considered by introducing a noise model on the preference while simultaneously relaxing the assumption of context independence to derive a sample complexity analysis for this more general setting.

A final crucial aspect concerns the determination of the context vectors. In our study, they were assumed to be provided by an oracle, however, given the availability of two response proposals, it is possible to estimate the corresponding context values using established NLP techniques. This would make the approach more autonomous.

Using these extensions, we anticipate that the proposed methodology could produce more accurate and generalizable results, thus improving its applicability to complex real-world environments.

Bibliography

[1] Yanqing Duan, John S Edwards, and Yogesh K Dwivedi. Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda. *International journal of information management*, 48:63–71, 2019.

[2] Constantin Hubmann, Marvin Becker, Daniel Althoff, David Lenz, and Christoph Stiller. Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles. In *2017 IEEE intelligent vehicles symposium (IV)*, pages 1671–1678. IEEE, 2017.

[3] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[4] Eric Topol. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

[5] Ernie Chan. *Algorithmic trading: winning strategies and their rationale*. John Wiley & Sons, 2013.

[6] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

[7] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.

[8] Thorsten Brants, Ashok Popat, Peng Xu, Franz Josef Och, and Jeffrey Dean. Large language models in machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 858–867, 2007.

[9] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B Hashimoto. Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, 12:39–57, 2024.

[10] Konstantinos I Roumeliotis and Nikolaos D Tselikas. Chatgpt and open-ai models: A preliminary review. *Future Internet*, 15(6):192, 2023.

[11] Keivalya Pandya and Mehfuza Holia. Automating customer service using langchain: Building custom open-source gpt chatbot for organizations. *arXiv preprint arXiv:2310.05421*, 2023.

[12] Marco Cascella, Jonathan Montomoli, Valentina Bellini, and Elena Bignami. Evaluating the feasibility of chatgpt in healthcare: an analysis of multiple clinical and research scenarios. *Journal of medical systems*, 47(1):33, 2023.

[13] Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2:79–84, 2021.

[14] Xinyi Hou, Yanjie Zhao, Yue Liu, Zhou Yang, Kailong Wang, Li Li, Xiapu Luo, David Lo, John Grundy, and Haoyu Wang. Large language models for software engineering: A systematic literature review. *ACM Transactions on Software Engineering and Methodology*, 33(8):1–79, 2024.

[15] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.

[16] Yufan Zhou, Ruiyi Zhang, Changyou Chen, Chunyuan Li, Chris Tensmeyer, Tong Yu, Jiuxiang Gu, Jinhui Xu, and Tong Sun. Towards language-free training for text-to-image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17907–17917, 2022.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[18] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[19] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? pages 610–623, 2021.

[20] Ralph L Keeney and Howard Raiffa. *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge university press, 1993.

[21] Edmundas Kazimieras Zavadskas, Zenonas Turskis, and Simona Kildienė. State of art surveys of overviews on mcdm/madm methods. *Technological and economic development of economy*, 20(1):165–179, 2014.

[22] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual review of psychology*, 62(2011):451–482, 2011.

[23] Blanca Ceballos, María Teresa Lamata, and David A Pelta. A comparative analysis of multi-criteria decision-making methods. *Progress in Artificial Intelligence*, 5:315–322, 2016.

[24] Thomas L Saaty. The analytic hierarchy process (ahp). *The Journal of the Operational Research Society*, 41(11):1073–1076, 1980.

[25] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 20(2):215–232, 1958.

[26] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[27] John W Chinneck. *Feasibility and Infeasibility in Optimization:: Algorithms and Computational Methods*, volume 118. Springer Science & Business Media, 2007.

[28] Kimia Ghobadi and Houra Mahmoudzadeh. Inferring linear feasible regions using inverse optimization. *European Journal of Operational Research*, 290(3):829–843, 2021.

[29] George B Dantzig. Origins of the simplex method. In *A history of scientific computing*, pages 141–151. 1990.

[30] James Renegar. *A mathematical view of interior-point methods in convex optimization.* SIAM, 2001.

[31] Haesol Im and Henry Wolkowicz. Revisiting degeneracy, strict feasibility, stability, in linear programming. *European Journal of Operational Research*, 310(2):495–510, 2023.

[32] Margaret Wright. The interior-point revolution in optimization: history, recent developments, and lasting consequences. *Bulletin of the American mathematical society*, 42(1):39–56, 2005.

[33] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[34] Jan Leike, David Krueger, Tom Everitt, Miljan Martic, Vishal Maini, and Shane Legg. Scalable agent alignment via reward modeling: a research direction. *arXiv preprint arXiv:1811.07871*, 2018.

[35] Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

[36] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[37] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

[38] Jian Hu, Li Tao, June Yang, and Chandler Zhou. Aligning language models with offline learning from human feedback. *arXiv preprint arXiv:2308.12050*, 2023.

[39] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36:53728–53741, 2023.

[40] Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, et al. Open problems and fundamental limitations of reinforcement learning from human feedback. *arXiv preprint arXiv:2307.15217*, 2023.

[41] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[42] Yvonne M Bishop, Stephen E Fienberg, and Paul W Holland. *Discrete multivariate analysis: Theory and practice*. Springer Science & Business Media, 2007.

[43] Morris H DeGroot. Probability and statistics. 1986.

[44] PennState. Stat 414: Introduction to probability theory. 2020.

[45] A. Cigna. Arctan of ratio of two normal variables is uniform. 2020.

[46] Walter Rudin. *Principles of mathematical analysis.* 2021.

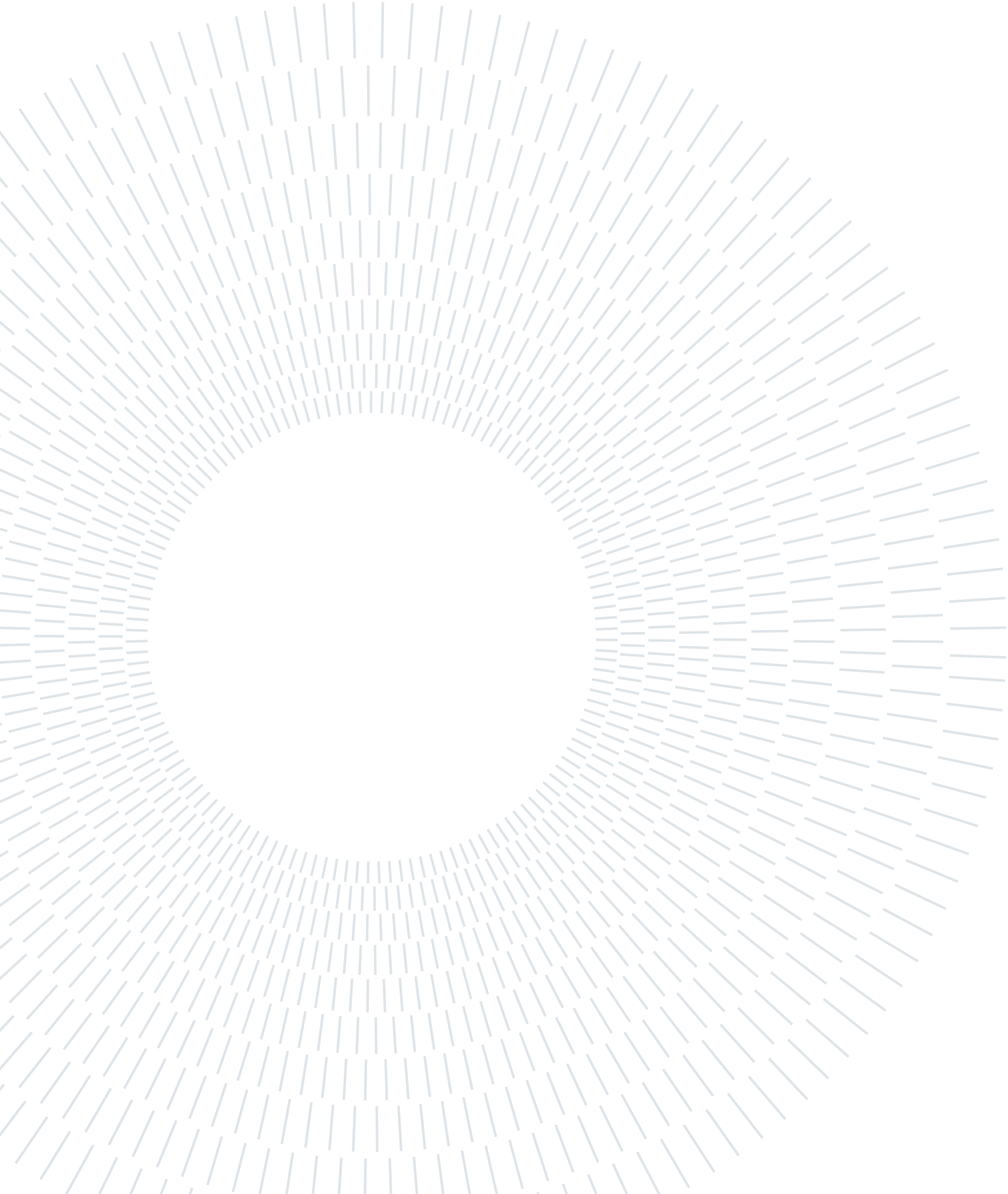
[47] O. Holder. *Ueber einen Mittelwertsatz.* 1889.

[48] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.

[49] Carl-Gustav Esseen. *A moment inequality with an application to the central limit theorem*, volume 1956. Taylor & Francis, 1956.

[50] Junlong Li, Fan Zhou, Shichao Sun, Yikai Zhang, Hai Zhao, and Pengfei Liu. Dissecting human and llm preferences. *arXiv preprint arXiv:2402.11296*, 2024.

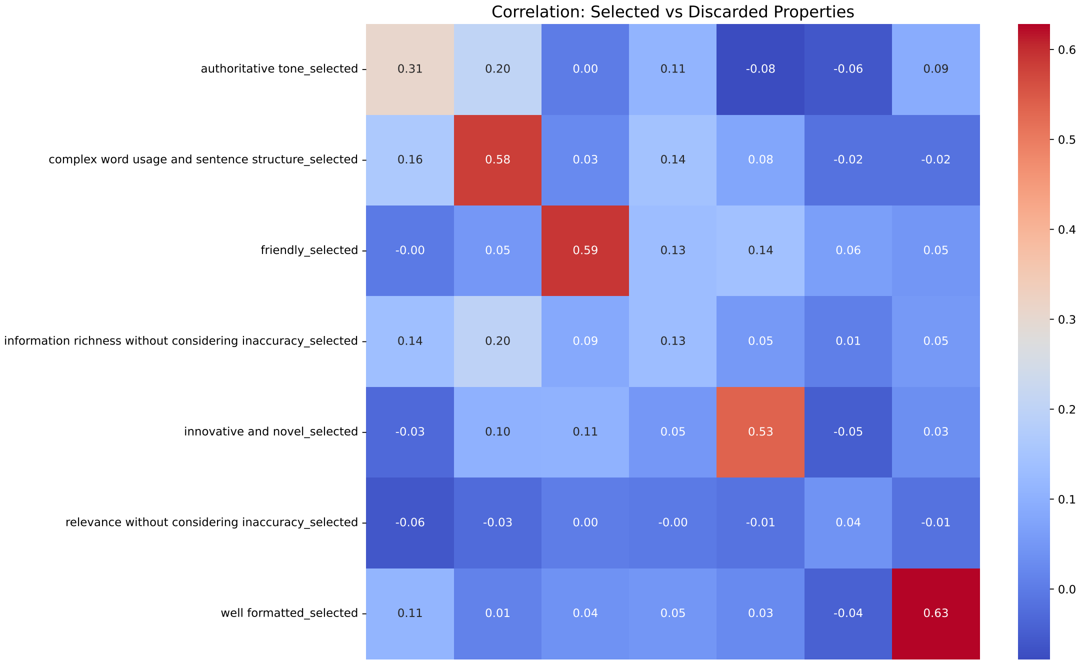
[51] G. Gordon and R. Tibshirani. Optimization. 2012.



A

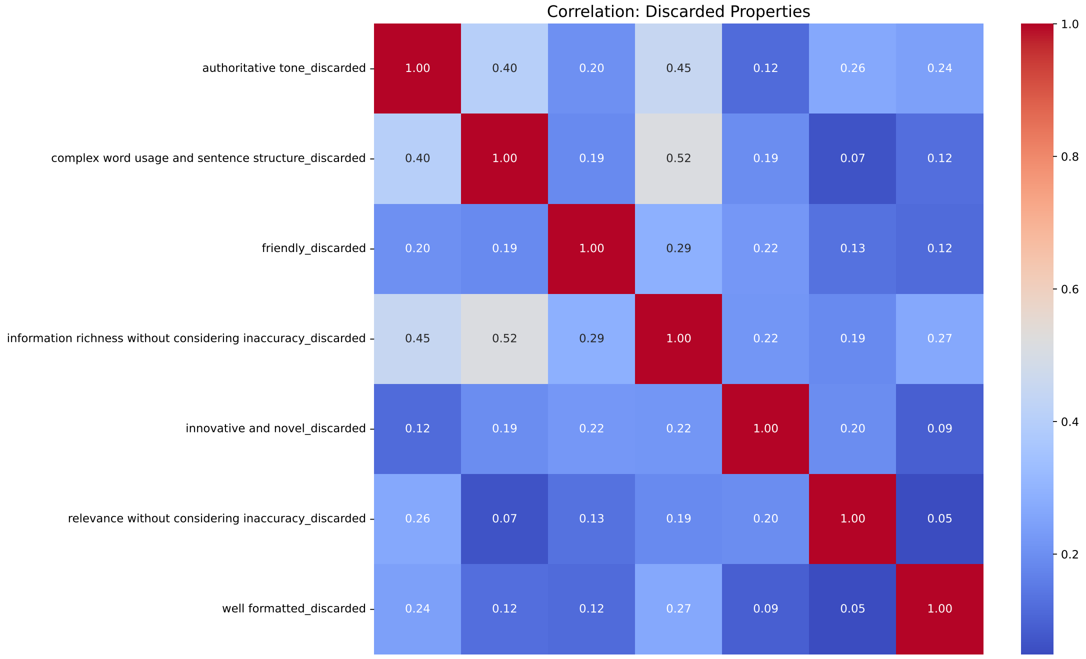
Appendix Contexts Selection

In this Appendix, we report more details on the context selection process. Figure A.1 shows the correlation matrices of the 7 chosen contexts divided respectively into the vectors of the selected responses, those of the discarded ones and comparing the former against the latter. As can be seen, the contexts that were chosen all have quite low correlation.

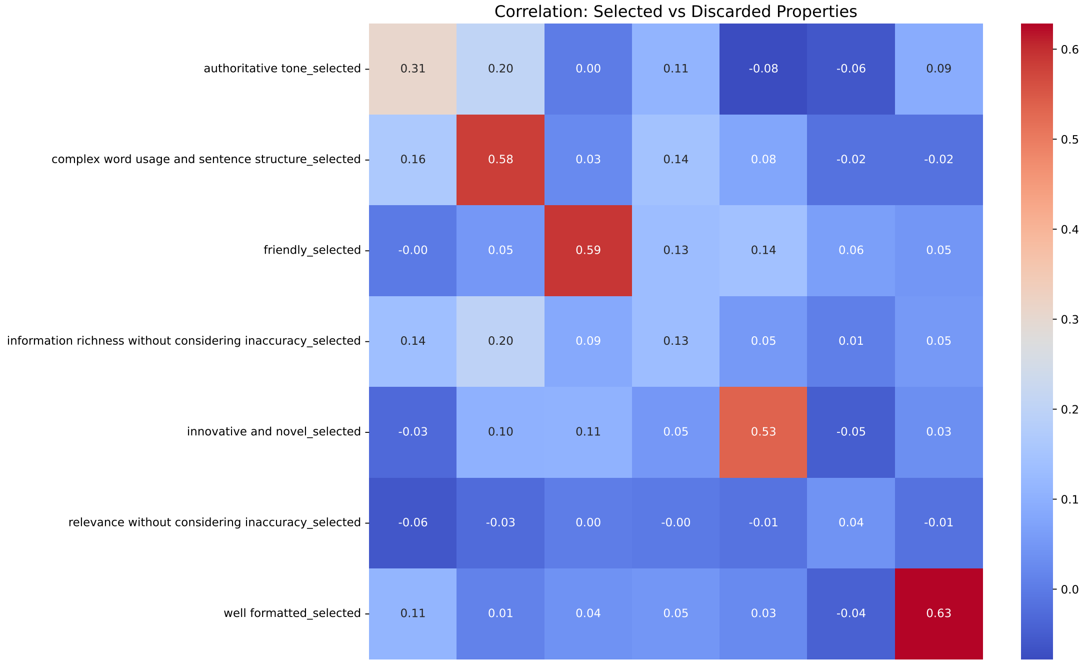


authoritative tone_selected	0.31	0.20	0.00	0.11	-0.08	-0.06	0.09
complex word usage and sentence structure_selected	0.16	0.58	0.03	0.14	0.08	-0.02	-0.02
friendly_selected	-0.00	0.05	0.59	0.13	0.14	0.06	0.05
information richness without considering inaccuracy_selected	0.14	0.20	0.09	0.13	0.05	0.01	0.05
innovative and novel_selected	-0.03	0.10	0.11	0.05	0.53	-0.05	0.03
relevance without considering inaccuracy_selected	-0.06	-0.03	0.00	-0.00	-0.01	0.04	-0.01
well formatted_selected	0.11	0.01	0.04	0.05	0.03	-0.04	0.63

(a) Selected.



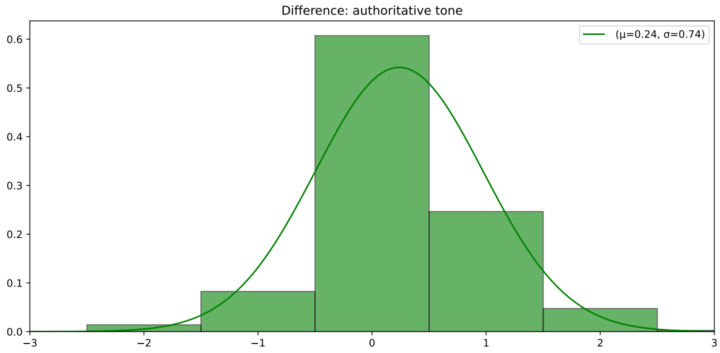
(b) Discarded.



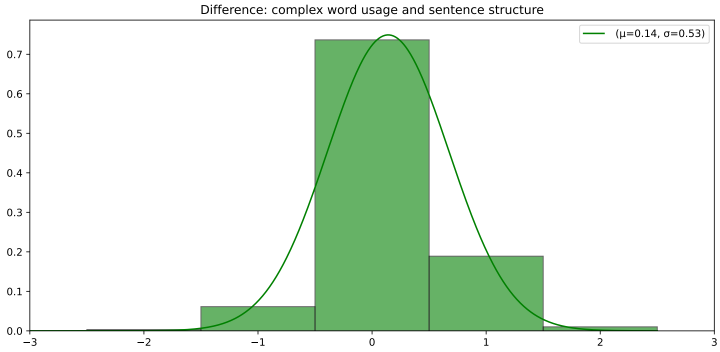
(c) Selected vs Discarded.

Figure A.1: Visualization of the correlation matrix of the 7 categories chosen.

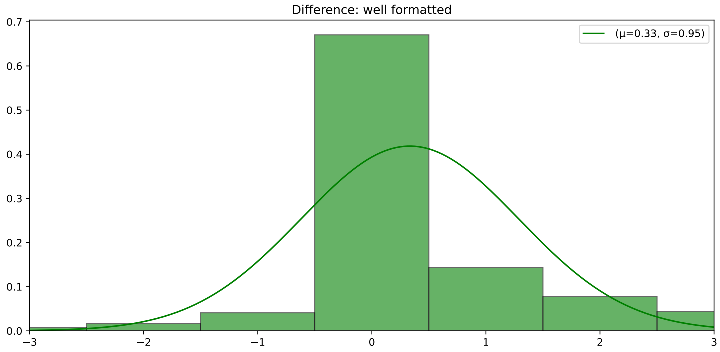
Figures A.2, A.3 and A.4 show the histograms with the different values that the contexts have in the dataset used and how the different distributions fit the vectors of real contexts. Notably, we can also observe differences in the values from the vectors relating to the selected answers compared to those of the discarded answers.



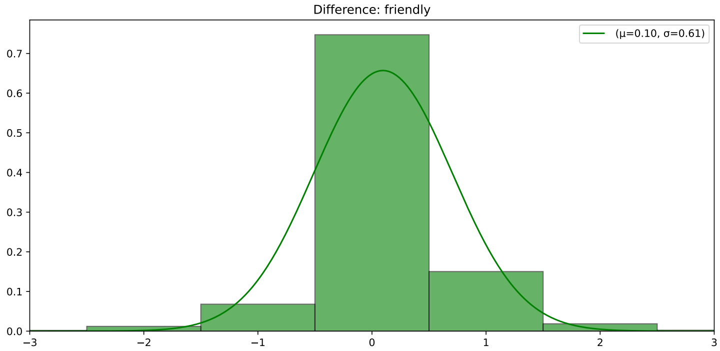
(a) Authoritative Tone.



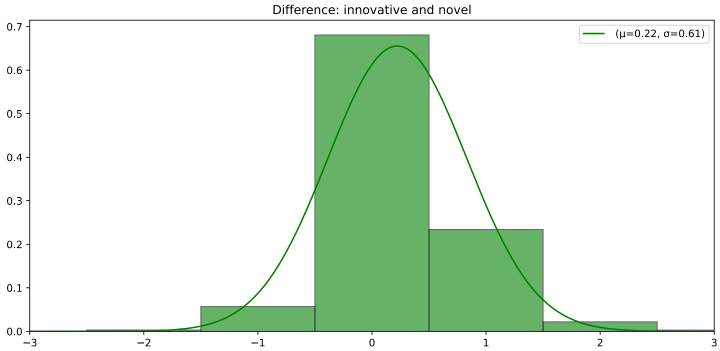
(b) Complex Word Usage.



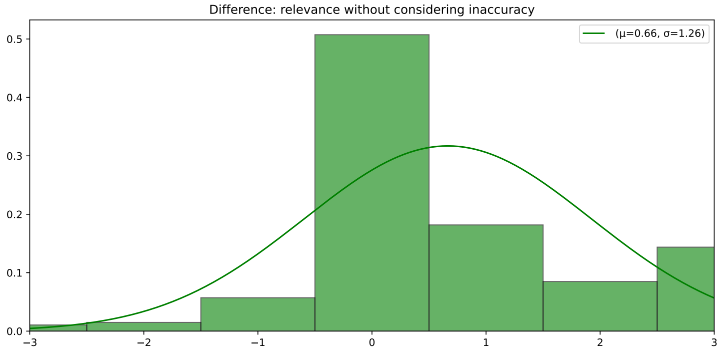
(c) Well Formatted.



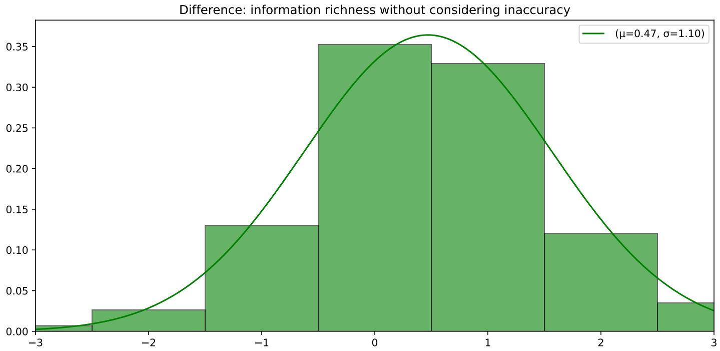
(d) Friendly.



(e) Innovative and Novel.

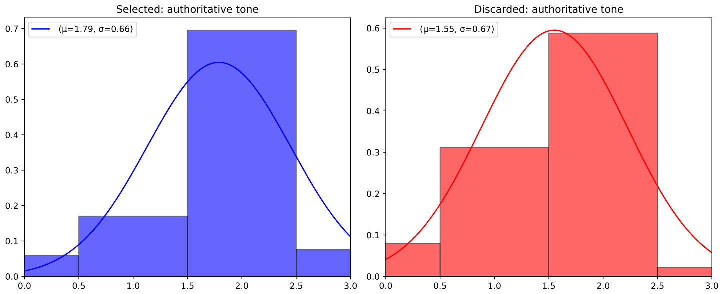


(f) Relevance without Inaccuracy.

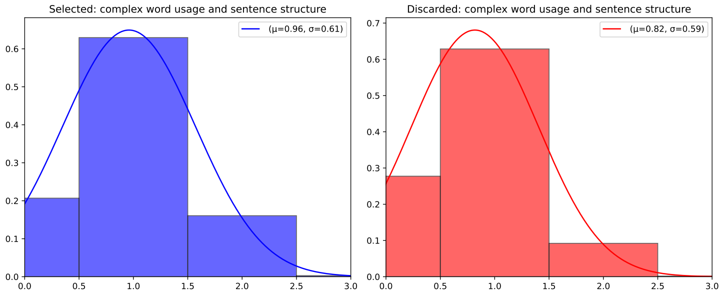


(g) Information Richness.

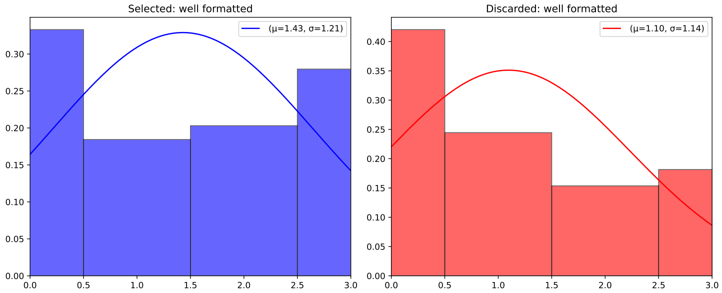
Figure A.2: Fit of a Gaussian on the difference $\mathbf{c}_{\text{selected}}$ and $\mathbf{c}_{\text{discarded}}$ of the categories chosen.



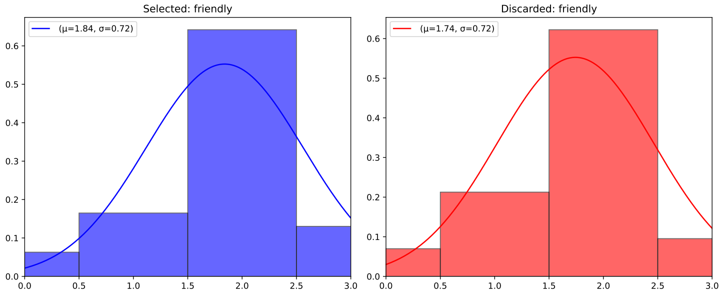
(a) Authoritative Tone.



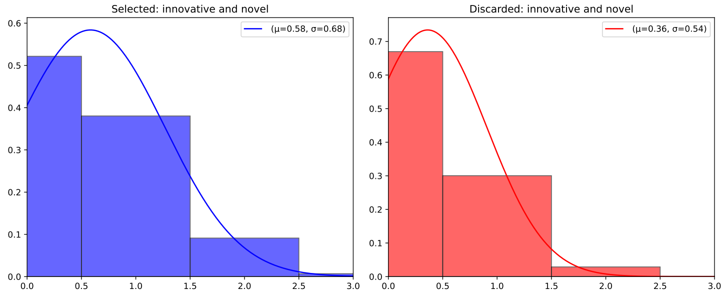
(b) Complex Word Usage.



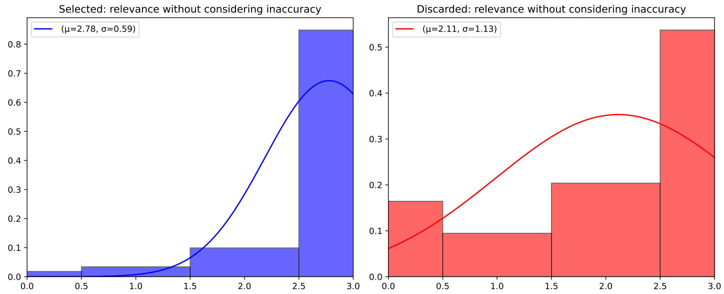
(c) Well Formatted.



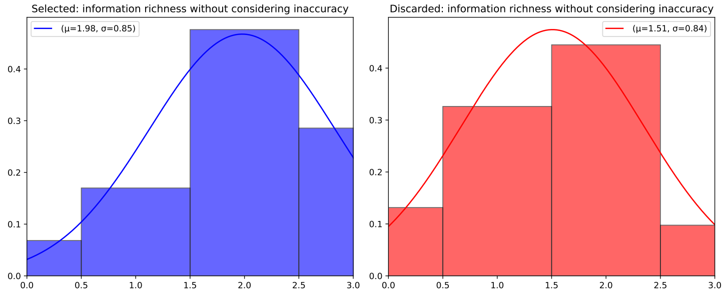
(d) Friendly.



(e) Innovative and Novel.

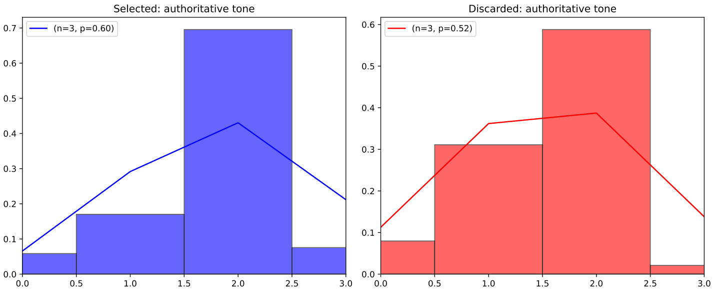


(f) Relevance without Inaccuracy.

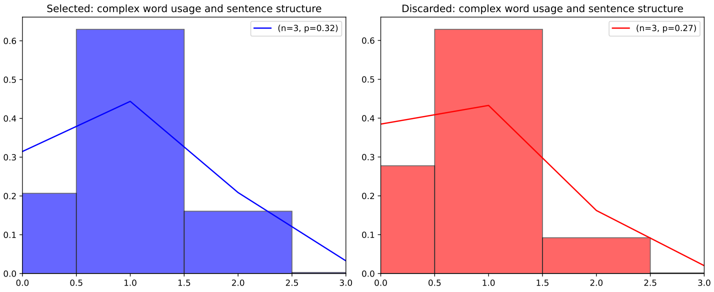


(g) Information Richness.

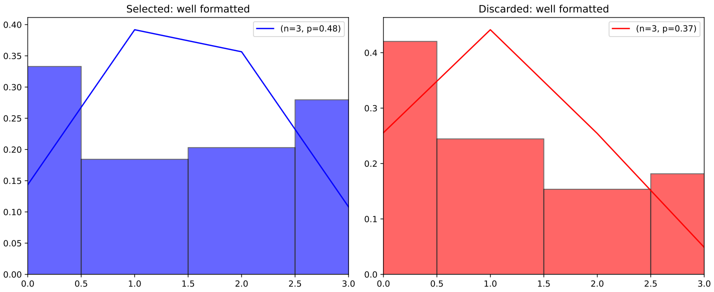
Figure A.3: Fit of a Gaussian on $\mathbf{c}_{selected}$ (in blue) and $\mathbf{c}_{discarded}$ (in red) of the categories chosen.



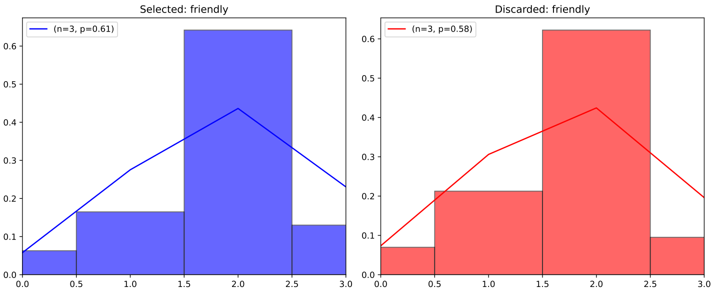
(a) Authoritative Tone.



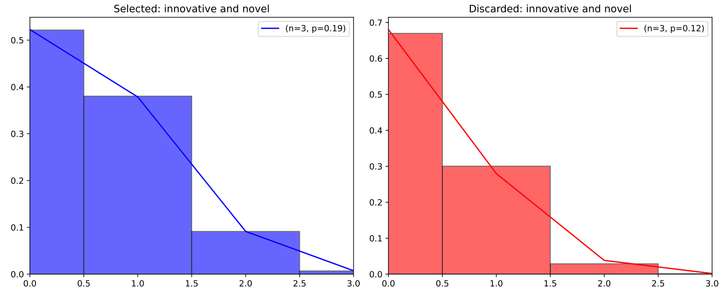
(b) Complex Word Usage.



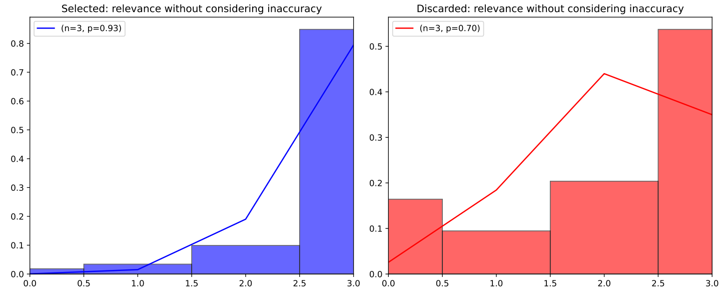
(c) Well Formatted.



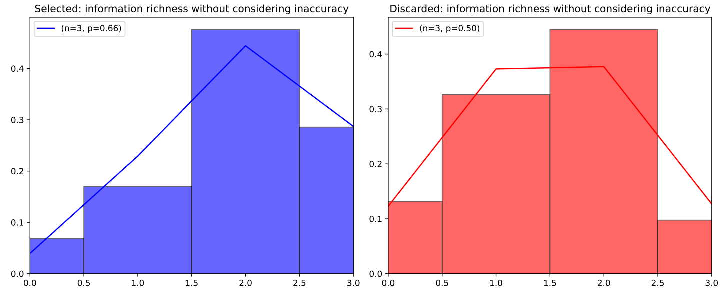
(d) Friendly.



(e) Innovative and Novel.



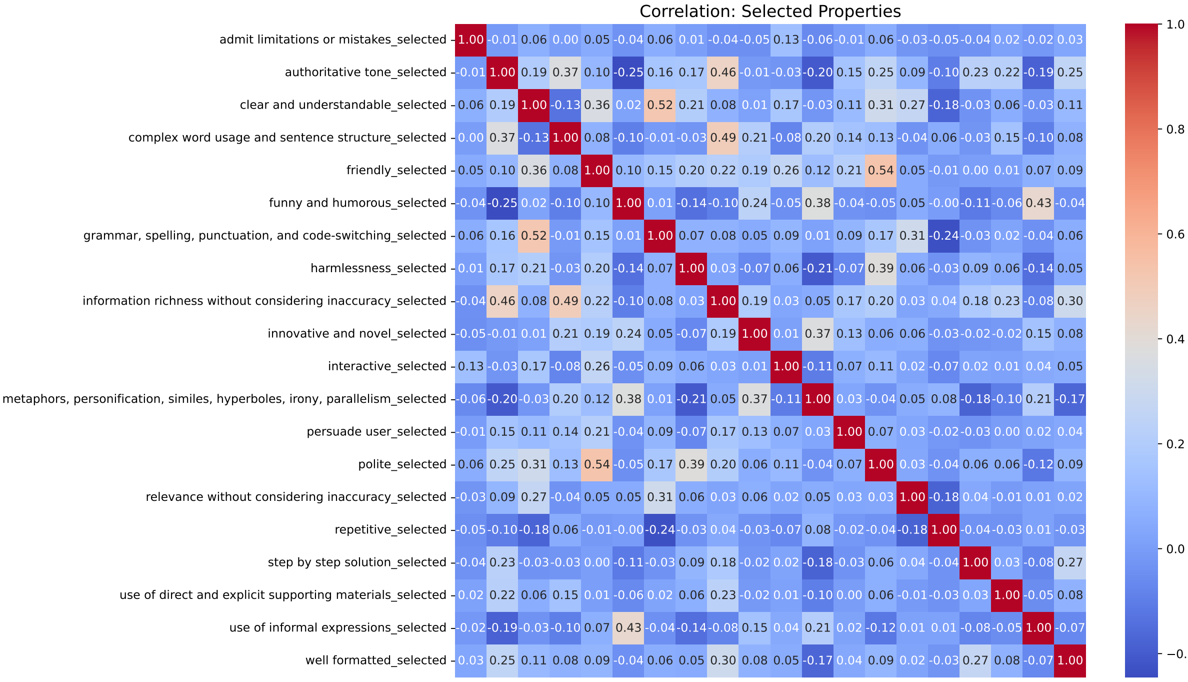
(f) Relevance without Inaccuracy.



(g) Information Richness.

Figure A.4: Fit of a Binomial on $\mathbf{c}_{selected}$ (in blue) and $\mathbf{c}_{discarded}$ (in red) of the categories chosen.

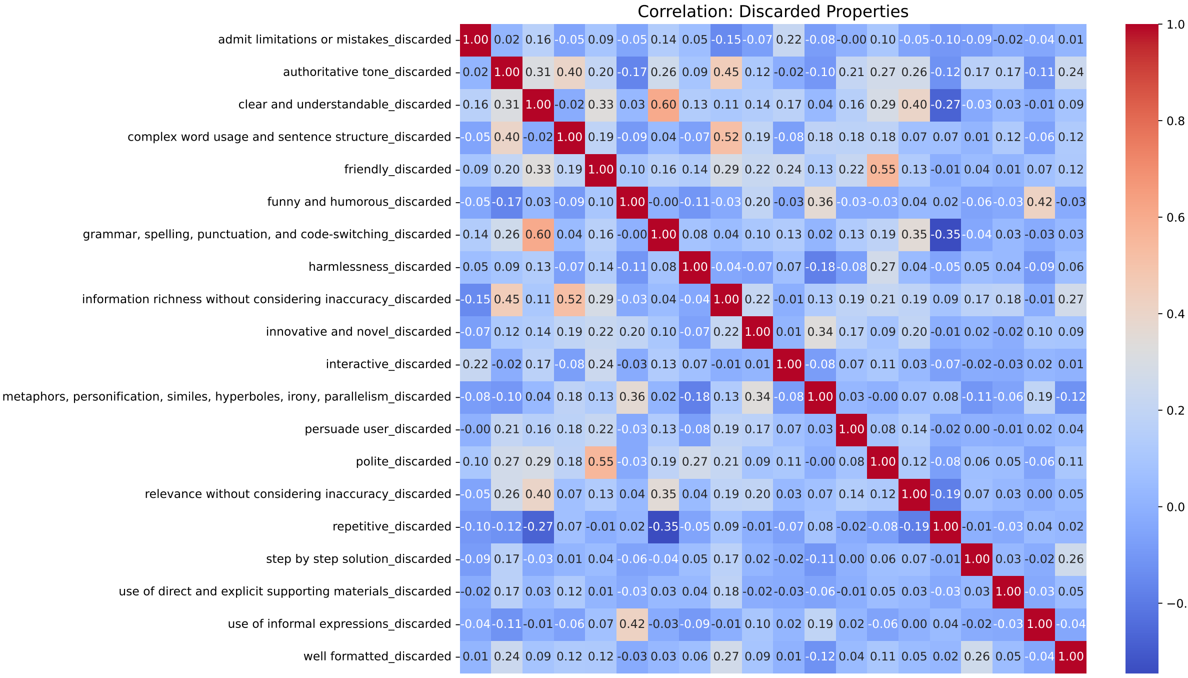
Finally, for completeness, in Figure A.5 we also report the correlation matrices relating to all 20 contexts present in the original dataset.



Correlation: Selected Properties

admit limitations or mistakes_selected	1.00	-0.01	0.06	0.00	0.05	-0.04	0.06	0.01	-0.04	-0.05	0.13	-0.06	-0.01	0.06	-0.03	-0.05	-0.04	0.02	-0.02	0.03
authoritative tone_selected	-0.01	1.00	0.19	0.37	0.10	-0.25	0.16	0.17	0.46	-0.01	-0.03	-0.20	0.15	0.25	0.09	-0.10	0.23	0.22	-0.19	0.25
clear and understandable_selected	0.06	0.19	1.00	-0.13	0.36	0.02	0.52	0.21	0.08	0.01	0.17	-0.03	0.11	0.31	0.27	-0.18	-0.03	0.06	-0.03	0.11
complex word usage and sentence structure_selected	0.00	0.37	-0.13	1.00	0.08	-0.10	-0.01	-0.03	0.49	0.21	-0.08	0.20	0.14	0.13	-0.04	0.06	-0.03	0.15	-0.10	0.08
friendly_selected	0.05	0.10	0.36	0.08	1.00	0.10	0.15	0.20	0.22	0.19	0.26	0.12	0.21	0.54	0.05	-0.01	0.00	0.01	0.07	0.09
funny and humorous_selected	-0.04	-0.25	0.02	-0.10	0.10	1.00	0.01	-0.14	-0.10	0.24	-0.05	0.38	-0.04	-0.05	0.05	-0.00	-0.11	-0.06	0.43	-0.04
grammar, spelling, punctuation, and code-switching_selected	0.06	0.16	0.52	-0.01	0.15	0.01	1.00	0.07	0.08	0.05	0.09	0.01	0.09	0.17	0.31	-0.24	-0.03	0.02	-0.04	0.06
harmlessness_selected	0.01	0.17	0.21	-0.03	0.20	-0.14	0.07	1.00	0.03	-0.07	0.06	-0.21	-0.07	0.39	0.06	-0.03	0.09	0.06	-0.14	0.05
information richness without considering inaccuracy_selected	-0.04	0.46	0.08	0.49	0.22	-0.10	0.08	0.03	1.00	0.19	0.03	0.05	0.17	0.20	0.03	0.04	0.18	0.23	-0.08	0.30
innovative and novel_selected	-0.05	-0.01	0.01	0.21	0.19	0.24	0.05	-0.07	0.19	1.00	0.01	0.37	0.13	0.06	0.06	-0.03	-0.02	-0.02	0.15	0.08
interactive_selected	0.13	-0.03	0.17	-0.08	0.26	-0.05	0.09	0.06	0.03	0.01	1.00	-0.11	0.07	0.11	0.02	-0.07	0.02	0.01	0.04	0.05
metaphors, personification, similes, hyperboles, irony, parallelism_selected	-0.06	-0.20	-0.03	0.20	0.12	0.38	0.01	-0.21	0.05	0.37	-0.11	1.00	0.03	-0.04	0.05	0.08	-0.18	-0.10	0.21	-0.17
persuade user_selected	-0.01	0.15	0.11	0.14	0.21	-0.04	0.09	-0.07	0.17	0.13	0.07	0.03	1.00	0.07	0.03	-0.02	-0.03	0.00	0.02	0.04
polite_selected	0.06	0.25	0.31	0.13	0.54	-0.05	0.17	0.39	0.20	0.06	0.11	-0.04	0.07	1.00	0.03	-0.04	0.06	0.06	-0.12	0.09
relevance without considering inaccuracy_selected	-0.03	0.09	0.27	-0.04	0.05	0.05	0.31	0.06	0.03	0.06	0.02	0.05	0.03	0.03	1.00	-0.18	0.04	-0.01	0.01	0.02
repetitive_selected	-0.05	-0.10	-0.18	0.06	-0.01	-0.00	-0.24	-0.03	0.04	-0.03	-0.07	0.08	-0.02	-0.04	-0.18	1.00	-0.04	-0.03	0.01	-0.03
step by step solution_selected	-0.04	0.23	-0.03	-0.03	0.00	-0.11	-0.03	0.09	0.18	-0.02	0.02	-0.18	-0.03	0.06	0.04	-0.04	1.00	0.03	-0.08	0.27
use of direct and explicit supporting materials_selected	0.02	0.22	0.06	0.15	0.01	-0.06	0.02	0.06	0.23	-0.02	0.01	-0.10	0.00	0.06	-0.01	-0.03	0.03	1.00	-0.05	0.08
use of informal expressions_selected	-0.02	-0.19	-0.03	-0.10	0.07	0.43	-0.04	-0.14	-0.08	0.15	0.04	0.21	0.02	-0.12	0.01	0.01	-0.08	-0.05	1.00	-0.07
well formatted_selected	0.03	0.25	0.11	0.08	0.09	-0.04	0.06	0.05	0.30	0.08	0.05	-0.17	0.04	0.09	0.02	-0.03	0.27	0.08	-0.07	1.00

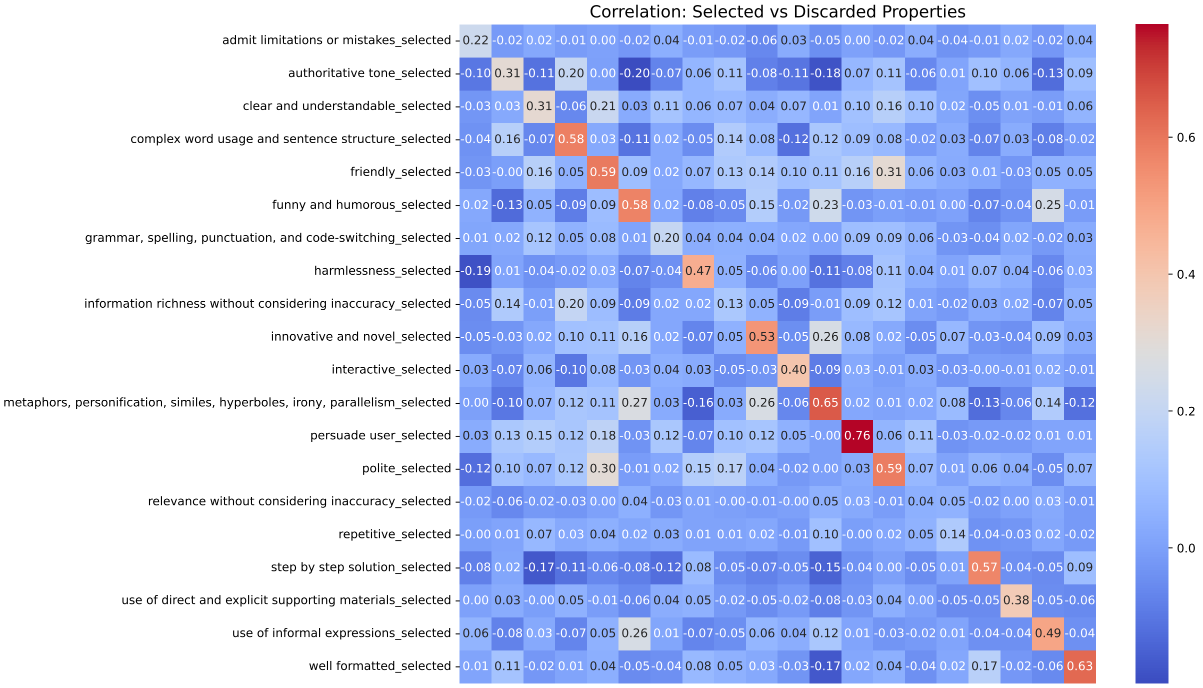
(a) Selected.



Correlation: Discarded Properties

admit limitations or mistakes_discarded	1.00	0.02	0.16	-0.05	0.09	-0.05	0.14	0.05	-0.15	-0.07	0.22	-0.08	-0.00	0.10	-0.05	-0.10	-0.09	-0.02	-0.04	0.01
authoritative tone_discarded	0.02	1.00	0.31	0.40	0.20	-0.17	0.26	0.09	0.45	0.12	-0.02	-0.10	0.21	0.27	0.26	-0.12	0.17	0.17	-0.11	0.24
clear and understandable_discarded	0.16	0.31	1.00	-0.02	0.33	0.03	0.60	0.13	0.11	0.14	0.17	0.04	0.16	0.29	0.40	-0.27	-0.03	0.03	-0.01	0.09
complex word usage and sentence structure_discarded	-0.05	0.40	-0.02	1.00	0.19	-0.09	0.04	-0.07	0.52	0.19	-0.08	0.18	0.18	0.18	0.07	0.07	0.01	0.12	-0.06	0.12
friendly_discarded	0.09	0.20	0.33	0.19	1.00	0.10	0.16	0.14	0.29	0.22	0.24	0.13	0.22	0.55	0.13	-0.01	0.04	0.01	0.07	0.12
funny and humorous_discarded	-0.05	-0.17	0.03	-0.09	0.10	1.00	-0.00	-0.11	-0.03	0.20	-0.03	0.36	-0.03	-0.03	0.04	0.02	-0.06	-0.03	0.42	-0.03
grammar, spelling, punctuation, and code-switching_discarded	0.14	0.26	0.60	0.04	0.16	-0.00	1.00	0.08	0.04	0.10	0.13	0.02	0.13	0.19	0.35	-0.35	-0.04	0.03	-0.03	0.03
harmlessness_discarded	0.05	0.09	0.13	-0.07	0.14	-0.11	0.08	1.00	-0.04	-0.07	0.07	-0.18	-0.08	0.27	0.04	-0.05	0.05	0.04	-0.09	0.06
information richness without considering inaccuracy_discarded	-0.15	0.45	0.11	0.52	0.29	-0.03	0.04	-0.04	1.00	0.22	-0.01	0.13	0.19	0.21	0.19	0.09	0.17	0.18	-0.01	0.27
innovative and novel_discarded	-0.07	0.12	0.14	0.19	0.22	0.20	0.10	-0.07	0.22	1.00	0.01	0.34	0.17	0.09	0.20	-0.01	0.02	-0.02	0.10	0.09
interactive_discarded	0.22	-0.02	0.17	-0.08	0.24	-0.03	0.13	0.07	-0.01	0.01	1.00	-0.08	0.07	0.11	0.03	-0.07	-0.02	-0.03	0.02	0.01
metaphors, personification, similes, hyperboles, irony, parallelism_discarded	-0.08	-0.10	0.04	0.18	0.13	0.36	0.02	-0.18	0.13	0.34	-0.08	1.00	0.03	-0.00	0.07	0.08	-0.11	-0.06	0.19	-0.12
persuade user_discarded	-0.00	0.21	0.16	0.18	0.22	-0.03	0.13	-0.08	0.19	0.17	0.07	0.03	1.00	0.08	0.14	-0.02	0.00	-0.01	0.02	0.04
polite_discarded	0.10	0.27	0.29	0.18	0.55	-0.03	0.19	0.27	0.21	0.09	0.11	-0.00	0.08	1.00	0.12	-0.08	0.06	0.05	-0.06	0.11
relevance without considering inaccuracy_discarded	-0.05	0.26	0.40	0.07	0.13	0.04	0.35	0.04	0.19	0.20	0.03	0.07	0.14	0.12	1.00	-0.19	0.07	0.03	0.00	0.05
repetitive_discarded	-0.10	-0.12	-0.27	0.07	-0.01	0.02	-0.35	-0.05	0.09	-0.01	-0.07	0.08	-0.02	-0.08	-0.19	1.00	-0.01	-0.03	0.04	0.02
step by step solution_discarded	-0.09	0.17	-0.03	0.01	0.04	-0.06	-0.04	0.05	0.17	0.02	-0.02	-0.11	0.00	0.06	0.07	-0.01	1.00	0.03	-0.02	0.26
use of direct and explicit supporting materials_discarded	-0.02	0.17	0.03	0.12	0.01	-0.03	0.03	0.04	0.18	-0.02	-0.03	-0.06	-0.01	0.05	0.03	-0.03	0.03	1.00	-0.03	0.05
use of informal expressions_discarded	-0.04	-0.11	-0.01	-0.06	0.07	0.42	-0.03	-0.09	-0.01	0.10	0.02	0.19	0.02	-0.06	0.00	0.04	-0.02	-0.03	1.00	-0.04
well formatted_discarded	0.01	0.24	0.09	0.12	0.12	-0.03	0.03	0.06	0.27	0.09	0.01	-0.12	0.04	0.11	0.05	0.02	0.26	0.05	-0.04	1.00

(b) Discarded.



admit limitations or mistakes_selected	0.22	-0.02	0.02	-0.01	0.00	-0.02	0.04	-0.01	-0.02	-0.06	0.03	-0.05	0.00	-0.02	0.04	-0.04	-0.01	0.02	-0.02	0.04
authoritative tone_selected	-0.10	0.31	-0.11	0.20	0.00	-0.20	-0.07	0.06	0.11	-0.08	-0.11	-0.18	0.07	0.11	-0.06	0.01	0.10	0.06	-0.13	0.09
clear and understandable_selected	-0.03	0.03	0.31	-0.06	0.21	0.03	0.11	0.06	0.07	0.04	0.07	0.01	0.10	0.16	0.10	0.02	-0.05	0.01	-0.01	0.06
complex word usage and sentence structure_selected	-0.04	0.16	-0.07	0.58	0.03	-0.11	0.02	-0.05	0.14	0.08	-0.12	0.12	0.09	0.08	-0.02	0.03	-0.07	0.03	-0.08	-0.02
friendly_selected	-0.03	-0.00	0.16	0.05	0.59	0.09	0.02	0.07	0.13	0.14	0.10	0.11	0.16	0.31	0.06	0.03	0.01	-0.03	0.05	0.05
funny and humorous_selected	0.02	-0.13	0.05	-0.09	0.09	0.58	0.02	-0.08	-0.05	0.15	-0.02	0.23	-0.03	-0.01	-0.01	0.00	-0.07	-0.04	0.25	-0.01
grammar, spelling, punctuation, and code-switching_selected	0.01	0.02	0.12	0.05	0.08	0.01	0.20	0.04	0.04	0.04	0.02	0.00	0.09	0.09	0.06	-0.03	-0.04	0.02	-0.02	0.03
harmlessness_selected	-0.19	0.01	-0.04	-0.02	0.03	-0.07	-0.04	0.47	0.05	-0.06	0.00	-0.11	-0.08	0.11	0.04	0.01	0.07	0.04	-0.06	0.03
information richness without considering inaccuracy_selected	-0.05	0.14	-0.01	0.20	0.09	-0.09	0.02	0.02	0.13	0.05	-0.09	-0.01	0.09	0.12	0.01	-0.02	0.03	0.02	-0.07	0.05
innovative and novel_selected	-0.05	-0.03	0.02	0.10	0.11	0.16	0.02	-0.07	0.05	0.53	-0.05	0.26	0.08	0.02	-0.05	0.07	-0.03	-0.04	0.09	0.03
interactive_selected	0.03	-0.07	0.06	-0.10	0.08	-0.03	0.04	0.03	-0.05	-0.03	0.40	-0.09	0.03	-0.01	0.03	-0.03	-0.00	-0.01	0.02	-0.01
metaphors, personification, similes, hyperboles, irony, parallelism_selected	0.00	-0.10	0.07	0.12	0.11	0.27	0.03	-0.16	0.03	0.26	-0.06	0.65	0.02	0.01	0.02	0.08	-0.13	-0.06	0.14	-0.12
persuade user_selected	0.03	0.13	0.15	0.12	0.18	-0.03	0.12	-0.07	0.10	0.12	0.05	-0.00	0.76	0.06	0.11	-0.03	-0.02	-0.02	0.01	0.01
polite_selected	-0.12	0.10	0.07	0.12	0.30	-0.01	0.02	0.15	0.17	0.04	-0.02	0.00	0.03	0.59	0.07	0.01	0.06	0.04	-0.05	0.07
relevance without considering inaccuracy_selected	-0.02	-0.06	-0.02	-0.03	0.00	0.04	-0.03	0.01	-0.00	-0.01	-0.00	0.05	0.03	-0.01	0.04	0.05	-0.02	0.00	0.03	-0.01
repetitive_selected	-0.00	0.01	0.07	0.03	0.04	0.02	0.03	0.01	0.01	0.02	-0.01	0.10	-0.00	0.02	0.05	0.14	-0.04	-0.03	0.02	-0.02
step by step solution_selected	-0.08	0.02	-0.17	-0.11	-0.06	-0.08	-0.12	0.08	-0.05	-0.07	-0.05	-0.15	-0.04	0.00	-0.05	0.01	0.57	-0.04	-0.05	0.09
use of direct and explicit supporting materials_selected	0.00	0.03	-0.00	0.05	-0.01	-0.06	0.04	0.05	-0.02	-0.05	-0.02	-0.08	-0.03	0.04	0.00	-0.05	-0.05	0.38	-0.05	-0.06
use of informal expressions_selected	0.06	-0.08	0.03	-0.07	0.05	0.26	0.01	-0.07	-0.05	0.06	0.04	0.12	0.01	-0.03	-0.02	0.01	-0.04	-0.04	0.49	-0.04
well formatted_selected	0.01	0.11	-0.02	0.01	0.04	-0.05	-0.04	0.08	0.05	0.03	-0.03	-0.17	0.02	0.04	-0.04	0.02	0.17	-0.02	-0.06	0.63

(c) Selected vs Discarded.

Figure A.5: Visualization of the correlation matrix of the 20 categories.

List of Figures

List of Tables