Theoretical Guarantees for Minimum Bayes Risk Decoding

Anonymous ACL submission

Abstract

Minimum Bayes Risk (MBR) decoding optimizes output selection by maximizing the expected utility value of an underlying human distribution. While prior work has shown the effectiveness of MBR decoding through empirical evaluation, few studies have analytically investigated why the method is effective. As a result of our analysis, we show that, given the size n of the reference hypothesis set used in computation, MBR decoding approaches the optimal solution with high probability at a rate of $\mathcal{O}\left(n^{-\frac{1}{2}}\right)$, under certain assumptions, even though the language space $\mathcal{Y}$ is significantly larger $\mathcal{Y} \gg n$. This result helps to theoretically explain the strong performance observed in several prior empirical studies on MBR decoding. In addition, we provide the performance gap for maximum-a-posteriori (MAP) decoding and compare it to MBR decoding. The result of this paper indicates that MBR decoding tends to converge to the optimal solution faster than MAP decoding in several cases.

1 Introduction

Minimum Bayes Risk (MBR) decoding (Kumar and Byrne, 2002, 2004) is a decision rule used to generate sequences from autoregressive probability models (e.g., LLMs). MBR decoding has been shown to produce high-quality texts in various directed text generation tasks, such as machine translation (Tromble et al., 2008; de Gispert et al., 2009; Stahlberg et al., 2017), text summarization (Rush et al., 2015; Narayan et al., 2018), text simplification (Heineman et al., 2024), image captioning (Borgeaud and Emerson, 2020), and instruction-following (Wu et al., 2025). Numerous experiments have reported the advantages of MBR decoding over maximum-a-posteriori (MAP) decoding (e.g., beam search) (Ehling et al., 2007; Eikema and Aziz, 2020; Müller and Sennrich, 2021; Eikema and Aziz, 2022; Bertsch et al., 2023).

Experimental results confirm that the larger the number of candidates and hypothesis sets collected, the better performance (Eikema and Aziz, 2022; Freitag et al., 2022). However, there is no theoretical explanation for the convergence rate of approaching optimal output. The answers to this question are the number of elements in the candidate and the hypothesis set in this paper. Our results show the following theorem.

Theorem. *(Convergence Rate of MBR Decoding; Informal) Under certain assumptions, MBR decoding approaches the optimal solution with high probability at a rate of* $\mathcal{O}\left(n^{-\frac{1}{2}}\right)$ *for the size* n *of the reference hypothesis set.*

This theoretical result is consistent with the empirical results of previous studies (Eikema and Aziz, 2022; Freitag et al., 2022). We also confirm that if the human distribution is similar to the model distribution, the performance of MBR decoding can be improved, as indicated by Ohashi et al. (2024). In addition, we derive the convergence rate of the optimal output for MAP decoding and compare it to MBR decoding. Our results show that MBR decoding tends to converge faster than MAP decoding in several cases.

Specifically, our main contributions are that we provide high probability and expected regret's upper bounds by MBR decoding in several cases (Theorem 1, Theorem 2, and Corollary 1) and we compare the performance gap and convergence rate of MBR decoding and MAP decoding within the same framework of the upper bound we derived in Section 5.

In summary, there are few theoretical analyses of MBR decoding, and thus a comprehensive theoretical framework has yet to be fully established. Through these contributions, we believe that this

study offers new perspectives that advance the understanding of MBR decoding.

2 Background and Notations

Text generation involves producing an output sequence based on an input sequence, the set of input sequences is defined by $\mathcal{X}$. Probabilistic text generators define a probability distribution over the output space of hypotheses $\mathcal{Y}$. The set of complete hypotheses $\mathcal{Y}$ is:

$$\mathcal{Y} := \{\text{BOS} \circ \boldsymbol{v} \circ \text{EOS} | \boldsymbol{v} \in \mathcal{V}^*\}.$$

where $\circ$ is a string concatenation and $\mathcal{V}^*$ is the Kleene closure of a set of vocabulary $\mathcal{V}$. The goal of decoding is to find the best hypothesis for a given input. For simplicity we write $\mathcal{M}_{\mathcal{Y}}^{\mathcal{X}}$ to denote a set of conditional probability distributions over a finite set $\mathcal{Y}$, given $\mathcal{X}$ as context sets. and $\mathcal{O}(n)$ is Big O notation.

2.1 MBR Decoding

Let $\mathcal{X}$ denote the input space and $\mathcal{Y}$ the output space. Given an input $x \in \mathcal{X}$, a probabilistic model defines a distribution $p \in \mathcal{M}_{\mathcal{Y}}^{\mathcal{X}}$ over possible outputs $y' \in \mathcal{Y}$. The goal of Bayes Risk minimization in structured prediction and sequence generation tasks is to select an output that minimizes the expected loss relative to the true distribution (Bach, 2024).

For a loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, the Bayes Risk is defined as:

$$\mathcal{R}(y \mid x) = \mathbb{E}_p \left[\ell(y', y)\right].$$

$$y^* = \arg\min_{y \in \mathcal{Y}} \mathcal{R}(y \mid x)$$

If the goal is to maximize some utility function u rather than to minimize a loss, it can also be interpreted as a performance metric $\ell = -u$.

The objective of MBR decoding is similar to the Bayes Risk, finding the output that maximizes the expected utility, thereby effectively minimizing risk (Kumar and Byrne, 2002, 2004).

The procedure consists of two key components: the human distribution $P_{\text{human}} \in \mathcal{M}_{\mathcal{Y}}^{\mathcal{X}}$ given input $x \in \mathcal{X}$ and a utility function. For simplicity, let $P(y \mid x) = P(y)$, since x is fixed in this paper. The utility function evaluates the quality of a candidate output $\mathcal{H} \subseteq \mathcal{Y}$ with respect to a reference output $\mathcal{Y}$. In this paper, we assume that the candidate hypotheses $\mathcal{H}$ are identical to the reference outputs $\mathcal{Y}$. Ideally, MBR decoding selects the optimal hypothesis by maximizing its expected utility over the distribution of human references:

$$u_h(y) = \sum_{y' \in \mathcal{Y}} u(y, y') \cdot P_{\text{human}}(y'). \quad (1)$$

$$y^* = \arg\max_{y \in \mathcal{Y}} u_h(y). \quad (2)$$

where utility function u: $\mathcal{Y} \times \mathcal{Y} \to [0, U_{\max}]$, $U_{\max} \in [0, 1]$ denotes the maximum utility value.

Since P_{human} is unknown, MBR decoding instead uses $P_{\text{model}} \in \mathcal{M}_{\mathcal{Y}}^{\mathcal{X}}$ to approximate P_{human}.

$$u_m(y) = \sum_{y' \in \mathcal{Y}} u(y, y') \cdot P_{\text{model}}(y'). \quad (3)$$

$$y^{\text{m}} = \arg\max_{y \in \mathcal{Y}} u_m(y). \quad (4)$$

However, summation over $\mathcal{Y}$ is computationally intractable, so Eq. (4) is approximated by a Monte Carlo estimation (Eikema and Aziz, 2022; Farinhas et al., 2023) using a collection of reference hypotheses $\mathcal{Y}_{\text{ref}}^{\text{n}}$ sampled from the model P_{model}.

$$\widehat{u}_m(y) = \frac{1}{|\mathcal{Y}_{\text{ref}}^{\text{n}}|} \sum_{y' \in \mathcal{Y}_{\text{ref}}^{\text{n}}} u(y, y'). \quad (5)$$

$$\hat{y}^m = \arg\max_{y \in \mathcal{Y}_{\text{ref}}^{\text{n}}} \widehat{u}_m(y). \quad (6)$$

We denote the number of samples used for the Monte Carlo estimate of the MBR decoding as $n := |\mathcal{Y}_{\text{ref}}^{\text{n}}|$.

Therefore, to derive a practical application equation (Eq. 6), two approximation operations are performed from the objective equation for true MBR decoding (Eq. 2).

2.2 MAP Decoding

The most intuitive decoding method is MAP decoding, which selects a mode based on the human distribution P_{human}. MAP decoding is also a special case in which the utility function of MBR decoding is used as the indicator function. The objective function of MAP decoding is defined by:

$$y^*_{\text{MAP}} = \arg\max_{y \in \mathcal{Y}} P_{\text{human}}(y). \quad (7)$$

The objective equation using the model probability P_{model} is similar to the MBR decoding:

$$y^m_{\text{MAP}} = \arg\max_{y \in \mathcal{Y}} P_{\text{model}}(y). \quad (8)$$

In addition, the objective equation for the Monte Carlo estimation of the MAP decoding is defined as:

$$\widehat{P}(y) = \frac{\sum_{y' \in \mathcal{Y}_{\text{ref}}^{\text{n}}} \mathbb{I}\left(y = y'\right)}{\left|\mathcal{Y}_{\text{ref}}^{\text{n}}\right|}. \tag{9}$$

where $\mathcal{Y}_{\text{ref}}^{\text{n}}$ collected n samples from P_{model}.

We reformulate the practical objective function of MAP decoding using Eq. (9):

$$\hat{y}_{\text{MAP}}^{m} = \arg\max_{y \in \mathcal{Y}_{\text{ref}}^{\text{n}}} \widehat{P}(y). \tag{10}$$

Eq. (10) shows the computationally feasible approximation of the MAP decoding. While beam search is the most common sampling strategy to approximate MAP decoding.

We focus on the analysis of MAP decoding and MBR decoding with random sampling in this paper, the case of considering temperature sampling for MBR decoding in Appendix G. The goal of the study is to investigate the statistics of the MBR and MAP objectives.

3 Analysis of MBR decoding

In this section, we evaluate the performance of MBR decoding (Eq. 6) compared to the ideal MBR solution (Eq. 2) under various assumptions.

3.1 Problem Setting

The optimal MBR decoding output is y^* (Eq. 2). However, as mentioned earlier, due to practical limitations, only a Monte Carlo solution $\hat{y}^m$ (Eq. 6) can be obtained in practice. On the other hand, since this $\hat{y}^m$ is ultimately evaluated under the human distribution P_{human}, the following performance difference arises:

$$\text{Regret}_n \coloneqq u_h(y^*) - u_h(\hat{y}^m). \tag{11}$$

We refer to this quantity Regret_n as regret. The goal of this study is to obtain an upper bound of regret theoretically. Suppose we can show this upper bound on the order of the number of elements in candidate and hypothesis sets. In that case, we can provide a theoretical guarantee for the performance of MBR decoding using the Monte Carlo method.

3.2 The Analysis of Regret_n

We define the following notation:

$$\Delta(u_p, u_q, y) \coloneqq u_p(y) - u_q(y). \tag{12}$$

This expresses the residual of the utility of y assuming q as the probability distribution when the target probability distribution is p. Using Eq. (12), we divide the regret Regret_n into four terms:

$$\begin{aligned}\text{Regret}_n \leq\ & \Delta(u_h, u_m, y^*) + \Delta(u_m, \widehat{u}_m, y^*) \\ & + \Delta(\widehat{u}_m, u_m, \hat{y}^m) + \Delta(u_m, u_h, \hat{y}^m).\end{aligned} \tag{13}$$

In the following analysis, we first derive an upper bound for each Δ in Eq. (13). Then, using the upper bounds derived for each Δ, we derive an upper bound for Regret_n.

Analysis of u_m and $\widehat{u}_m$. First, we derive an upper bound for the terms involving $u_m, \widehat{u}_m$. We consider the following assumption about the utility function.

Assumption 1 (Inner Product Representation of the Utility Function)**.** *Let $\boldsymbol{\alpha}(y) \in \mathbb{R}^d$ be an embedding for each $y \in \mathcal{Y}$. We assume that the utility function $u(y, y')$ is given by the inner product of the embeddings, i.e.,*

$$u(y, y') = \boldsymbol{\alpha}(y)^\top \boldsymbol{\alpha}(y').$$

There are examples of utility functions that satisfy these properties such as the F_1 measure of the BERTScore and the inner product of the embedding functions (Zhang et al., 2020; Reimers and Gurevych, 2019; Meng et al., 2024). Note that several state-of-the-art utility functions for machine translation do not satisfy this assumption (e.g., COMET and Metric-X; Rei et al. 2020; Guerreiro et al. 2024; Juraska et al. 2024), since they are trained to do the computation of the utility and the quality estimation at the same time.

By applying Hoeffding's Inequality (Lemma 4) and Uniform Concentration Inequality (Lemma 5), see Appendix B, along with Assumption 1, the following lemma about the terms involving $u_m, \widehat{u}_m$ is established.

Lemma 1 (Upper Bound for the terms involving $u_m, \widehat{u}_m$)**.** *Under Assumption 1 and assuming $d \geq 4$, the following bound holds for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$\begin{aligned}&\Delta(u_m, \widehat{u}_m, y^*) + \Delta(\widehat{u}_m, u_m, \hat{y}^m) \\ &\leq 3\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d \log d}.\end{aligned}$$

The proof can be found in Appendix C. Since the dimensions of the embedding models are usually larger than 4, we assume them in this study and proceed with our analysis under this assumption.[1] Lemma 1 shows that the upper bound of regret with $u_m, \widehat{u}_m$ terms depends only on the number of samples n and decreases at a rate of $\mathcal{O}\left(n^{-\frac{1}{2}}\right)$. Notably, this result can also be interpreted as a regret bound, specifically Regret_n, under the condition that P_{human} and P_{model} are identical (Appendix D).

Analysis of u_h and u_m. Next, we analyze the u_h, u_m terms involved, but before doing so, we consider the following assumptions.

Assumption 2 (Utility Function Smoothness)**.** *For all $y, y', y'' \in \mathcal{Y}$, we assume the utility function satisfies the following inequality:*

$$|u(y, y') - u(y, y'')| \leq C(y', y'')$$

where $C \in \mathbb{R}^{\mathcal{Y} \times \mathcal{Y}}$ is a cost function.

The assumption is not a restrictive assumption. For any utility functions bounded by $[0, U_{\max}]$, $C(y', y'') = U_{\max}$ satisfies the assumption. Note also that Assumption 1 entails Assumption 2. The assumption is known as the Lipschitz condition (Jeffreys and Jeffreys, 1988). It claims that the value of the utility function is smooth under the cost function C: the difference in the utility between an output y and other outputs y' and y'' can be bounded by some "distance" C between the y' and y''. Intuitively, if y' and y'' are similar, then C wants to be small, and otherwise large. Many of the utility functions are designed to be so by minimizing the prediction error from the human evaluation (e.g., MQM score) (Rei et al., 2020; Juraska et al., 2024). Under the Assumption 2, the following lemma holds:

Lemma 2 (Upper Bound for the terms involving u_h, u_m)**.** *Under Assumption 2, the following bound can be derived:*

$$\Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, \hat{y}^m) \leq 2\mathrm{WD}(P_{\text{human}}, P_{\text{model}}),$$

where WD *is Wasseratein distance with C being the cost function.*

The definition of Wasserstein distance (Wang, 2012) is described in Appendix B. The proof can be found in Appendix E. Lemma 2 implies that minimizing the terms involving u_h, u_m requires choosing P_{model} that closely approximates P_{human}.

Upper bound of Regret_n. Using Lemma 1 and Lemma 2, we can derive an upper bound for Regret_n.

Theorem 1 (Regret Bound for MBR decoding)**.** *Under Assumption 1, Assumption 2, and assuming $d \geq 4$, the regret upper bound of the MBR decoding holds for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$\mathrm{Regret}_n \leq 3\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d} + 2\mathrm{WD}(P_{\text{human}}, P_{\text{model}}).$$

Theorem 1 can be interpreted as follows. First, the upper bound decreases with a larger number of samples from P_{model} with the convergence speed of $\mathcal{O}\left(n^{-\frac{1}{2}}\right)$. This implies that you can reduce the upper bound by 30% by doubling the number of samples $2n$, you are probably to need at least four times more samples $4n$ to reduce the initial error by 50%. The other insight is that the error is inherently limited by the Wasserstein distance between P_{human} and P_{model}, which means that, as expected, the accuracy of P_{model} is desirable. This observation is consistent with the finding that Ohashi et al. (2024) indicates that MBR decoding performance is improved when P_{human} and P_{model} are similar.

3.3 On the Effect of the Training Dataset Size

In practice, we cannot compute the exact value of the Wasserstein distance as it requires enumeration over all possible sentences. To derive a more digestible bound, we consider the simplest example where P_{model} is an empirical distribution of P_{human}. Formally, we consider the following assumption:

Assumption 3 (P_{model} as an Empirical Distribution Sampled the Size of Training

[1] For readers interested in the case $d < 4$, see Appendix C.

Dataset $|D|$ from P_{human}.). *Let P_{model} be the empirical distribution of $|D|$ samples obtained from P_{human}. P_{model} has the following expression:*

$$P_{\text{model}}(y) = \frac{1}{|D|} \sum_{y' \in D} \mathbb{I}(y = y').$$
$$D \sim P_{\text{human}}(\cdot)$$

where $\mathbb{I}$ is an indicator function.

Assumption 3 is not intended to be an assumption that is directly applicable to real-world scenarios. In a real-world scenario, P_{model} is almost always represented by function approximation models (e.g., neural networks) for text generation tasks. They are often pretrained by unsupervised learning using a language model objective, then fine-tuned by supervised learning and preference learning (Radford et al., 2018; Stiennon et al., 2020; Ouyang et al., 2022).

Given the diversity and complexity of models used in practice, we instead analyze a simple model where it has no function approximation, pertaining, or post-training processes. Such a simple model is likely to be worse than models used in practice. Therefore, the bounds we derive from this simple model serve as informal worst-case bounds for the state-of-the-art models.

The size of the training dataset is usually much larger than the number of samples for MBR decoding: $|D| \gg n$.

Analysis of u_m and $\widehat{u}_m$ with the training dataset size $|D|$. Under Assumption 3, we derive the analysis on the terms in u_h, u_m, using the Hoeffding's Inequality (Lemma 4, see Appendix B), we can get the following the upper bound.

Lemma 3 (Upper Bound for the terms involving u_h, u_m with the Size of Training Dataset $|D|$)**.** *Under Assumption 3, the following bound holds for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$\Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, \hat{y}^m) \leq 3\sqrt{\frac{1}{|D|} \log \frac{1}{\delta}}.$$

The proof can be found in Appendix F. This shows that the upper bounds for the u_h, u_m terms vary only with the size of the training dataset $|D|$ and that the upper bound decays at a rate of $\mathcal{O}\left(|D|^{-\frac{1}{2}}\right)$ with its size.

Under Assumption 3, regret depends on both samples n and $|D|$. For clarity, we define a regret under Assumption 3 as follows:

$$\text{Regret}_{n,D} \coloneqq u_h(y^*) - u_h(\hat{y}^m). \quad (14)$$

Upper bound of $\text{Regret}_{n,D}$. We can immediately derive the upper bound for $\text{Regret}_{n,D}$ from Lemma 1 and Lemma 3 as follows:

Theorem 2 (Regret Bound for MBR decoding with the Size of Training Dataset $|D|$)**.** *Under Assumption 1, Assumption 3, and assuming $d \geq 4$, the regret upper bound of the MBR decoding holds for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$\text{Regret}_{n,D} \leq 4\sqrt{\frac{1}{n} \log \frac{1}{\delta}} + 4\sqrt{\frac{1}{|D|} \log \frac{1}{\delta}} + \frac{36}{n}\sqrt{d \log d}.$$

Theorem 2 shows that MBR decoding approaches the optimal output with a convergence rate related to the size of the reference hypothesis set n and the size of the training dataset $|D|$, suggesting why MBR decoding has good experimental performance. This implies that sample and dataset sizes are significant for MBR decoding.

3.4 Extended Analysis of MBR Decoding

Expected regret bounds. So far, we have found that we can obtain upper bounds that occur with high probability, and from these upper bounds, we can immediately determine the expected regret upper bound for Theorem 1 and Theorem 2.

Corollary 1 (Expected Regret Upper Bound of MBR decoding)**.** *The expected regret upper bounds $\text{Regret}_n, \text{Regret}_{n,D}$ are bounded as follows for any $\delta \in (0, 1)$ under assuming $d \geq 4$:*

$$\mathbb{E}\left[\text{Regret}_n\right] \leq 3\sqrt{\frac{1}{n} \log \frac{1}{\delta}} + \frac{36}{n}\sqrt{d \log d} + 2\text{WD}(P_{\text{human}}, P_{\text{model}}) + \delta$$

$$\mathbb{E}\left[\mathrm{Regret}_{n,D}\right] \leq 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d} + 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + \delta$$

The proof is in Appendix H. Corollary 1 can be used to estimate how large the regret will be, on average.

On the effect of errors in the utility function. In the real world, we cannot always have access to the true utility function u, instead, we assume that the proxy utility function is only available u', and we explain the bound of the difference in the utility function. We focus exclusively on the conditions outlined in Theorem 2.

We define the following expression:

$$u'(y) = \frac{1}{|\mathcal{Y}^{\mathrm{n}}_{\mathrm{ref}}|}\sum_{y'\in\mathcal{Y}^{\mathrm{n}}_{\mathrm{ref}}} u'(y,y').$$
$$y' = \arg\max_{y\in\mathcal{Y}^{\mathrm{n}}_{\mathrm{ref}}} u'(y).$$

We want to find the upper bound of $u_h(y^*) - u_h(y')$. Our new objective function in this paragraph is defined as:

$$\mathrm{Regret}^{u}_{n,D} \coloneqq u_h(y^*) - u_h(y'). \tag{15}$$

Under Assumption 1 and Assumption 3, an upper bound of $\mathrm{Regret}^{u}_{n,D}$ is derived using the Hoeffding's inequality (Lemma 4). Let $\alpha_{err} := \max_{y,y'\in\mathcal{Y}^{\mathrm{n}}_{\mathrm{ref}}} ||\boldsymbol{\alpha}(y) - \boldsymbol{\alpha}'(y')||$.

Corollary 2 (Regret Bound for MBR decoding with utility function error)**.** *Under Assumption 1 and Assumption 3, the regret upper bound of the MBR decoding with utility function error holds for any $\delta \in (0,1)$, with probability at least $1-\delta$:*

$$\mathrm{Regret}^{u}_{n,D} \leq 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 2d\alpha_{err}.$$

The proof can be found in Appendix I. In Corollary 2, the upper bound decreases as the number of samples n and the size of training dataset $|D|$ increases. However, it does not ultimately converge to zero, as the term α_{err} remains.

4 Analysis of MAP Decoding

In this section, we derive the regret of MAP decoding between the optimal value and the Monte Carlo estimated value, expressed as $P_{\mathrm{human}}(y^*_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})$.

We define the regret of the MAP decoding as follows:

$$\mathrm{Regret}^{\mathrm{MAP}}_{n} = P_{\mathrm{human}}(y^*_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}). \tag{16}$$

We refer to $\mathrm{Regret}^{\mathrm{MAP}}_{n}$ as MAP regret. Under the conditions of Theorem 1, the upper bound of $\mathrm{Regret}^{\mathrm{MAP}}_{n}$ is given by the following result using Dvoretzky–Kiefer–Wolfowitz inequality (Lemma 6, see in Appendix B).

Theorem 3 (Regret Bound for MAP decoding)**.** *Under the conditions of Theorem 1, the MAP regret upper bound of the MAP decoding holds for any $\delta \in (0,1)$, with probability at least $1-\delta$:*

$$\mathrm{Regret}^{\mathrm{MAP}}_{n} \leq 6\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 2\mathrm{WD}(P_{\mathrm{human}}, P_{\mathrm{model}}).$$

Furthermore, under the conditions of Theorem 2, MAP regret depends on the number of samples n and the size of the training dataset $|D|$.

Our new objective formulation is defined as:

$$\mathrm{Regret}^{\mathrm{MAP}}_{n,D} = P_{\mathrm{human}}(y^*_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}) \tag{17}$$

The upper bound of $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$ is also immediately obtained by Lemma 6 as follows.

Theorem 4 (Regret Bound for MAP decoding with the Size of Training Dataset $|D|$)**.** *Under the conditions of Theorem 2, the MAP regret upper bound of the MAP decoding holds for any $\delta \in (0,1)$, with probability at least $1-\delta$:*

$$\mathrm{Regret}^{\mathrm{MAP}}_{n,D} \leq 8\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 8\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}.$$

The proof is in Appendix J. Note that Theorem 3 and Theorem 4 decrease in the same order Theorem 1 and Theorem 2 respectively. In other words, if we compare the difference between these bounds more clearly, we focus on the constant term.

5 Performance Comparison

So far, we have analyzed MBR decoding and MAP decoding independently. In this section, we compare the performance of MBR decoding and MAP decoding within the same framework, in terms of the upper bound and we focus exclusively on the conditions outlined in Theorem 2, Theorem 4.

Difference between MBR and MAP Decoding target values. First, we aim to analyze the difference $u_h(y^*) - u_h(\hat{y}^m_{\mathrm{MAP}})$, where y^* is the optimal output and $\hat{y}^m_{\mathrm{MAP}}$ is a suboptimal output. We can analyze this error bound to see how the optimal solution in MAP decoding behaves in an ideal MBR decoding environment.

Observation 1 (Error between y^* and $\hat{y}^m_{\mathrm{MAP}}$ with u_h)**.** *Error bound between y^* and $\hat{y}^m_{\mathrm{MAP}}$ with u_h satisfies for any $\delta \in \left(0, \frac{2}{5}\right)$, with probability at least $1 - \frac{5}{2}\delta$:*

$$u_h(y^*) - u_h(\hat{y}^m_{\mathrm{MAP}}) \leq u_m(\hat{y}^m) - u_m(\hat{y}^m_{\mathrm{MAP}}) + \mathcal{O}\left(\max\left(n^{-\frac{1}{2}}, |D|^{-\frac{1}{2}}\right)\right).$$

The detail is in Appendix K.1. This observation confirms that MAP decoding and MBR decoding theoretically have different objectives under certain conditions. We provide the analysis of the MAP regret $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$ of the two decoding algorithms in Appendix K.1.

Convergence speed of upper bound of $\mathrm{Regret}_{n,D}$ and $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$. Next, we compare the upper bounds of the convergence rate between MBR decoding and MAP decoding presented in this study.

Observation 2 (Comparison of the Convergence Speed)**.** *We compare the upper bounds of $\mathrm{Regret}_{n,D}$ and $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$ under three different scenarios:*

1. $n \to \infty$ and $|D|$ is finite. *The upper bound of $\mathrm{Regret}_{n,D}$ is strictly smaller than the upper bound of $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$.*

2. $D \to \infty$ and n is finite. *For number of samples n and utility d such that the following inequality holds, the upper bound of $\mathrm{Regret}_{n,D}$ is smaller than the upper bound of $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$.*

$$\frac{1}{9}\sqrt{n \log \frac{1}{\delta}} \geq \sqrt{d \log d}.$$

3. Both D and n are finite. *For the number of samples n, utility d, and the size of training dataset $|D|$ such that the following inequality holds, the upper bound of $\mathrm{Regret}_{n,D}$ is smaller than the upper bound of $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$.*

$$\frac{n}{9}\left(\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}\right) \geq \sqrt{d \log d}.$$

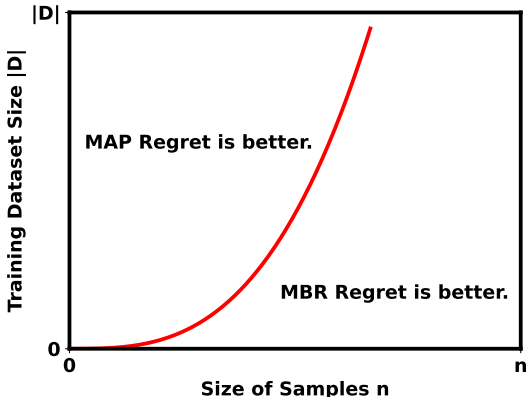

Figure 1: Conceptual visualization of Observation 2. The convergence rates of the upper bound of $\mathrm{Regret}_{n,D}$ and $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$ with the number of samples n and training dataset size $|D|$ are compared. For n and $|D|$ on the right side of this line plot, it means that the upper bound of $\mathrm{Regret}_{n,D}$ is smaller.

The proofs are given in Appendix K.2 and the result of numerical experiments comparing the rate of convergence of these upper bounds depending on the number of samples n is shown in Appendix L. As can be seen by comparing $\mathrm{Regret}_{n,D}$ and $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$ in cases 2 and 3 of Observation 2, as the number of samples n increases, the upper bound on $\mathrm{Regret}_{n,D}$ converges more faster (Fig. 1).

The analysis shows that, for any model, there exists a large enough n such that the upper bound of the regret of MBR decoding is smaller than that of MAP decoding. This observation may help explain why the empirical performance of MBR decoding can exceed that of MAP decoding.

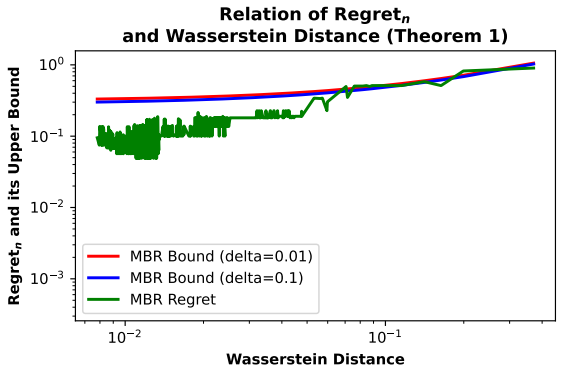

Figure 2: The upper bound ($\delta = \{0.01, 0.1\}$) of the Regret_n derived by Theorem 1 and the value of Regret_n in the simulation. The number of samples n is fixed to 500 and the training dataset size is $|D| = [0, 1000]$

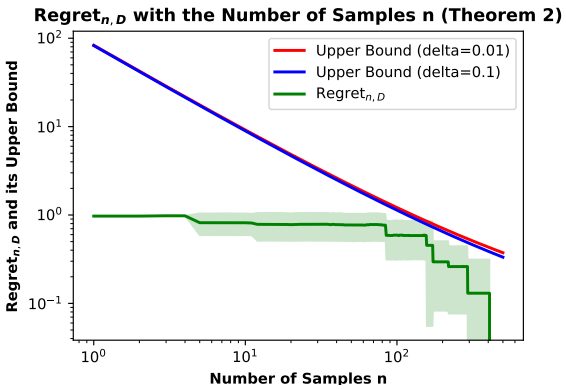

Figure 3: The upper bound ($\delta = \{0.01, 0.1\}$) of the $\mathrm{Regret}_{n,D}$ derived by Theorem 2 and the value of $\mathrm{Regret}_{n,D}$ in the simulation. The training dataset size $|D|$ is fixed to 5000 and the number of samples for MBR decoding is $n = [0, 500]$.

6 Numerical Simulation

In this section, we computationally evaluate the upper bounds of Regret_n and $\mathrm{Regret}_{n,D}$. It is important to emphasize that the experiments conducted in this paper are not intended to show the tightness of the results in the practical NLP tasks. Rather, they are intended to provide a visual representation of the theoretical results presented in this paper.

For the performance of MBR decoding in real-world NLP tasks, we refer to previous work (Freitag et al., 2023; Bertsch et al., 2023; Heineman et al., 2024; Wu et al., 2025).

We consider $\mathcal{Y} = 10{,}000$ hypotheses y_i ($i = 1, \ldots, \mathcal{H}$) with the dataset size of $|D|$ and the number of samples n model samples to study regret and bound behavior. We test $\delta \in \{0.01, 0.1\}$, and set $d = 4$. The details of the experimental setup are given in Appendix M.

6.1 Results

Fig. 2 clearly demonstrates that our theoretical upper bound on Regret_n is tight when compared to the actual regret observed. This close correspondence indicates that the assumptions and inequalities used in deriving the bound are well-justified, providing evidence in the numerical experiment.

The results of Fig. 3 and Fig. 4 show that the obtained upper bound converges to $\mathrm{Regret}_{n,D}$ as the number of samples increases. This behavior suggests the theoretical validity of the bound indicating that the looseness of the upper bound is gradually eliminated as the number of samples increases, improving the ability to accurately capture the true performance difference accurately.

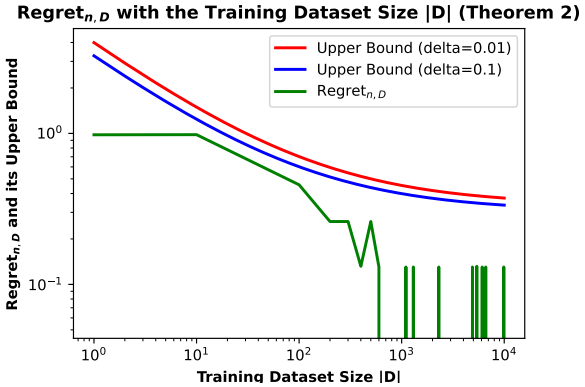

Figure 4: The upper bound ($\delta = \{0.01, 0.1\}$) of the $\mathrm{Regret}_{n,D}$ derived by Theorem 2 and the value of $\mathrm{Regret}_{n,D}$ in simulation. The number of samples n is fixed to 500 and the training dataset size is $|D| = [0, 10000]$

7 Conclusions

This paper presents a theoretical analysis of MBR decoding and shows that, under reasonable assumptions, it converges with high probability to the optimal solution at a rate of $\mathcal{O}\left(n^{-\frac{1}{2}}\right)$, even when the total language space $\mathcal{Y}$ is large. In addition, we compare MBR and MAP decoding about the performance difference and the convergence speed to the optimal solution. As a result, we confirm MAP decoding and MBR decoding theoretically have different objectives, and from the upper bound, MBR decoding is more efficient than MAP decoding in approaching the optimal output under certain conditions.

8 Limitations

This study provides the first theoretical bounds on MBR decoding. As it is one of the first analyses on MBR decoding, it has several limitations, particularly regarding its alignment with practical implementations.

Assumptions. Our analysis assumes that the set of candidate hypotheses $\mathcal{H}$ is identical to the set of reference outputs $\mathcal{Y}$. However, in practice, using a small number of high-quality but biased candidates alongside a larger set of unbiased references has been found to be more effective.

We have considered three assumptions for the analysis. The assumptions do not cover all the situations of text generation applications. For example, the state-of-the-art utility functions for machine translation (COMET and Metric-X; Rei et al. 2020; Guerreiro et al. 2024; Juraska et al. 2024) are not linear function (Assumption 1).

In practice, the models are represented by a neural network, and they are often pretrained using unsupervised learning before supervised fine-tuning. This point is not considered in Assumption 3.

Aspects not considered. The analysis does not account for the role of neural networks. In particular, it is known that solutions corresponding to flat minima tend to generalize better than those with sharp minima (Dinh et al., 2017). Understanding the role of neural networks for MBR decoding is future work.

We analyze a model that predicts sequences, but practical implementations typically use autoregressive language models (Lin et al., 2021). Incorporating the autoregressive assumption may lead to improved theoretical bounds.

The study considers only random sampling and temperature sampling. However, other strategies, such as epsilon sampling (Hewitt et al., 2022) and beam search (which is commonly used for MAP decoding), are not analyzed.

Our analysis does not frame the problem as an NLP task. Incorporating domain-specific characteristics could lead to tighter bounds. This study is purely theoretical and does not include empirical experiments to validate the results on real-world NLP tasks. Instead, we rely on prior experimental findings (Freitag et al., 2023; Bertsch et al., 2023; Suzgun et al., 2023) for providing empirical support for our theoretical conclusions.

Another key limitation is that the bounds derived in this study are not proven to be tight, leaving room for refinement. Furthermore, to measure how tight the upper bounds is, we also need to derive the lower bound in MBR decoding.

Lastly, while our study focuses on sample complexity, practical implementations of MBR decoding are often constrained by computational complexity (Cheng and Vlachos, 2023; Vamvas and Sennrich, 2024). Combining our sampling complexity result with the existing computational complexity bounds (Jinnai and Ariu, 2024) is future work.

Summary. This work provides fundamental theoretical bounds for MBR decoding. However, there remain avenues for improvement, including empirical validation, refinement of theoretical bounds, and comparative analysis with alternative decoding algorithms.

9 Ethics Statements

We do not foresee any ethical concerns regarding the analysis of the paper.

References

Francis Bach. 2024. *Learning Theory from First Principles*. Adaptive Computation and Machine Learning series. MIT Press.

Amanda Bertsch, Alex Xie, Graham Neubig, and Matthew Gormley. 2023. It's MBR all the way down: Modern generation techniques through the lens of minimum Bayes risk. In *Proceedings of the Big Picture Workshop*, pages 108–122, Singapore. Association for Computational Linguistics.

Sebastian Borgeaud and Guy Emerson. 2020. Leveraging sentence similarity in natural language generation: Improving beam search using range voting. In *Proceedings of the Fourth Workshop on Neural Generation and Translation*, pages 97–109, Online. Association for Computational Linguistics.

Julius Cheng and Andreas Vlachos. 2023. Faster minimum Bayes risk decoding with confidence-based pruning. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12473–12480, Singapore. Association for Computational Linguistics.

Adrià de Gispert, Sami Virpioja, Mikko Kurimo, and William Byrne. 2009. Minimum Bayes risk combination of translation hypotheses from alternative morphological decompositions. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion*

Volume: Short Papers, pages 73–76, Boulder, Colorado. Association for Computational Linguistics.

Laurent Dinh, Razvan Pascanu, Samy Bengio, and Yoshua Bengio. 2017. Sharp minima can generalize for deep nets. In *International Conference on Machine Learning*, pages 1019–1028. PMLR.

Nicola Ehling, Richard Zens, and Hermann Ney. 2007. Minimum Bayes risk decoding for BLEU. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 101–104, Prague, Czech Republic. Association for Computational Linguistics.

Bryan Eikema and Wilker Aziz. 2020. Is MAP decoding all you need? the inadequacy of the mode in neural machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4506–4520, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Bryan Eikema and Wilker Aziz. 2022. Sampling-Based Approximations to Minimum Bayes Risk Decoding for Neural Machine Translation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10978–10993, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

António Farinhas, José de Souza, and Andre Martins. 2023. An Empirical Study of Translation Hypothesis Ensembling with Large Language Models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 11956–11970, Singapore. Association for Computational Linguistics.

Patrick Fernandes, António Farinhas, Ricardo Rei, José De Souza, Perez Ogayo, Graham Neubig, and Andre Martins. 2022. Quality-aware decoding for neural machine translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1396–1412, Seattle, United States. Association for Computational Linguistics.

Markus Freitag, Behrooz Ghorbani, and Patrick Fernandes. 2023. Epsilon sampling rocks: Investigating sampling strategies for minimum Bayes risk decoding for machine translation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9198–9209, Singapore. Association for Computational Linguistics.

Markus Freitag, David Grangier, Qijun Tan, and Bowen Liang. 2022. High quality rather than high model probability: Minimum Bayes risk decoding with neural metrics. *Transactions of the Association for Computational Linguistics*, 10:811–825.

Nuno M. Guerreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André F. T. Martins. 2024. xcomet: Transparent machine translation evaluation through fine-grained error detection. *Transactions of the Association for Computational Linguistics*, 12:979–995.

David Heineman, Yao Dou, and Wei Xu. 2024. Improving Minimum Bayes Risk Decoding with Multi-Prompt. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22525–22545, Miami, Florida, USA. Association for Computational Linguistics.

John Hewitt, Christopher Manning, and Percy Liang. 2022. Truncation sampling as language model desmoothing. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3414–3427, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

H Jeffreys and BS Jeffreys. 1988. The Lipschitz Condition. *Methods of Mathematical Physics*, page 53.

Yuu Jinnai and Kaito Ariu. 2024. Hyperparameter-free approach for faster minimum Bayes risk decoding. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8547–8566, Bangkok, Thailand. Association for Computational Linguistics.

Juraj Juraska, Daniel Deutsch, Mara Finkelstein, and Markus Freitag. 2024. MetricX-24: The Google Submission to the WMT 2024 Metrics Shared Task. In *Proceedings of the Ninth Conference on Machine Translation*, pages 492–504, Miami, Florida, USA. Association for Computational Linguistics.

Hidetaka Kamigaito, Hiroyuki Deguchi, Yusuke Sakai, Katsuhiko Hayashi, and Taro Watanabe. 2024. Theoretical Aspects of Bias and Diversity in Minimum Bayes Risk Decoding. *arXiv preprint arXiv:2410.15021*.

Shankar Kumar and William Byrne. 2002. Minimum Bayes-risk word alignments of bilingual texts. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 140–147. Association for Computational Linguistics.

Shankar Kumar and William Byrne. 2004. Minimum Bayes-risk decoding for statistical machine translation. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 169–176, Boston, Massachusetts, USA. Association for Computational Linguistics.

Chu-Cheng Lin, Aaron Jaech, Xin Li, Matthew R. Gormley, and Jason Eisner. 2021. Limitations of autoregressive models and their alternatives. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5147–5173, Online. Association for Computational Linguistics.

Pascal Massart. 1990. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, pages 1269–1283.

Rui Meng, Ye Liu, Shafiq Rayhan Joty, Caiming Xiong, Yingbo Zhou, and Semih Yavuz. 2024. SFR-Embedding-2: Advanced Text Embedding with Multi-stage Training.

Mehryar Mohri. 2018. Foundations of machine learning.

Mathias Müller and Rico Sennrich. 2021. Understanding the properties of minimum Bayes risk decoding in neural machine translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 259–272, Online. Association for Computational Linguistics.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Atsumoto Ohashi, Ukyo Honda, Tetsuro Morimura, and Yuu Jinnai. 2024. On the True Distribution Approximation of Minimum Bayes-Risk Decoding. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 459–468, Mexico City, Mexico. Association for Computational Linguistics.

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

Gabriel Peyré and Marco Cuturi. 2020. Computational optimal transport. *arXiv preprint arXiv:1803.00567*.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.

Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Felix Stahlberg, Adrià de Gispert, Eva Hasler, and Bill Byrne. 2017. Neural machine translation by minimising the Bayes-risk with respect to syntactic translation lattices. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 362–368, Valencia, Spain. Association for Computational Linguistics.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021.

Mirac Suzgun, Luke Melas-Kyriazi, and Dan Jurafsky. 2023. Follow the wisdom of the crowd: Effective text generation via minimum Bayes risk decoding. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4265–4293, Toronto, Canada. Association for Computational Linguistics.

Roy Tromble, Shankar Kumar, Franz Och, and Wolfgang Macherey. 2008. Lattice Minimum Bayes-Risk decoding for statistical machine translation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 620–629, Honolulu, Hawaii. Association for Computational Linguistics.

Jannis Vamvas and Rico Sennrich. 2024. Linear-time minimum bayes risk decoding with reference aggregation. *arXiv preprint arXiv:2402.04251*.

Martin J Wainwright. 2019. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press.

Feng-Yu Wang. 2012. Coupling and applications. In *Stochastic Analysis and Applications to Finance: Essays in Honour of Jia-An Yan*, pages 411–424. World Scientific.

Ian Wu, Patrick Fernandes, Amanda Bertsch, Seungone Kim, Sina Khoshfetrat Pakazad, and Graham Neubig. 2025. Better Instruction-Following Through Minimum Bayes Risk. In *The Thirteenth International Conference on Learning Representations*.

Jianhao Yan, Jin Xu, Fandong Meng, Jie Zhou, and Yue Zhang. 2024. DC-MBR: Distributional Cooling for Minimum Bayesian Risk Decoding. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources*

and Evaluation (LREC-COLING 2024), pages 4423–4437, Torino, Italia. ELRA and ICCL.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

A Related Works

Experimental findings in MBR decoding. Many studies have reported that the performance of MBR decoding increases with a larger number of samples (Eikema and Aziz, 2022; Freitag et al., 2022). Prior works (Freitag et al., 2022; Fernandes et al., 2022) show that the performance of the MBR decoding depends on the selection of the utility function. Experiments combining MBR decoding with neural reference-based metrics, such as BLEURT, demonstrate significant improvements in human evaluations. In recent work, Yan et al. (2024) propose Distributional Cooling MBR, this approach bridges the gap between label smoothing and MBR decoding, with extensive experimental validation demonstrating its effectiveness on various NMT benchmarks and Wu et al. (2025) shows that leveraging reference-based LLM judges with MBR decoding improves the output quality of instruction-following LLMs compared to greedy decoding, best-of-N approaches.

Analysis of MBR decoding. Kamigaito et al. (2024) conduct on the intricate relationship between bias and diversity in MBR decoding. Their bias-diversity decomposition framework theoretically explains the trade-offs observed in empirical studies.

B Useful Lemmas and Definition

In this section, we present the concentration inequality used in the paper. The following inequalities represent a uniform concentration inequality.

Lemma 4 (Hoeffding's inequality; Corollary 1.1 in Bach 2024)**.** $\{X_i\}_{i=1}^n \in [0, b]$ *being i.i.d. samples drawn from same distribution.*

$$\Pr\left(\left|\mathbb{E}\left[X\right] - \frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{b^2}\right).$$

The following inequalities represent a uniform concentration inequality.

Lemma 5 (Uniform Concentration Inequality; Theorem 4.10 in Wainwright 2019)**.** $\mathcal{F}$ *is a class of functions* $f : \mathcal{X} \to [0, b]$.

$$\Pr\left(\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} \geq 2\mathcal{R}_n(\mathcal{F}) + \epsilon\right) \leq \exp\left(-\frac{2n\epsilon^2}{b^2}\right).$$

where $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{P}_n f - \mathbb{P}f|$, $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$ and $\mathbb{P}f = \mathbb{E}[f(X)]$, with X and $\{X_i\}_{i=1}^n$ being i.i.d. samples drawn from $\mathbb{P}$, $\mathcal{R}_n : (\mathcal{F}, \{X_i\}_{i=1}^n) \to \mathbb{R}$.

$\mathcal{R}_n(\mathcal{F})$ represents the Rademacher complexity of the function class $\mathcal{F}$ (Definition 3.1 (Mohri, 2018)). Rademacher complexity is a measure of model complexity, indicating how well a function class can fit random noise. It provides a uniform bound on the deviation between the empirical and true expectations across all functions in $\mathcal{F}$, serving as a key tool for analyzing generalization error in statistical learning theory.

Lemma 6 (Dvoretzky–Kiefer–Wolfowitz inequality; Massart 1990)**.**

$$\Pr\left(\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \varepsilon\right) \leq 2\exp(-2n\varepsilon^2).$$

Given a natural number n, let $X_1, X_2, \cdots, X_n$ be real-valued independent and identically distributed random variables with cumulative distribution function $F(\cdot)$. Let F_n denote the associated empirical distribution function defined by $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i\leq x\}}$

Definition 1 (Wassertstein Distance)**.** The Wassertstein Distance (WD) (Wang, 2012) is defined as:

$$\mathrm{WD}(\nu, \mu) = \inf_{\gamma\in\Gamma(\nu,\mu)} \sum_{(i,j)\in N\times N} \gamma_{ij}\, C_{ij}, \tag{18}$$

where N: the total number of samples, consisting of the set $\{y_1, y_2, \ldots, y_N\}$, $\nu, \mu \in \Delta_N$: probability measure on the aforementioned sets (ν_i, μ_i refer to the probability value $\nu(y_i), \mu(y_i)$), C: $N \times N \to \mathbb{R}$ a cost function measuring the distance between two outputs (e.g. C_{ij} refers to the amount to be transported from place y_i to palace y_j), and $\Gamma(\nu, \mu)$ denotes the set of all joint distributions γ whose marginals are ν and μ. The constraints on γ are given by:

$$\begin{aligned} \sum_{j \in n} \gamma_{ij} &= \nu_i, \quad \forall i \in n, \\ \sum_{i \in n} \gamma_{ij} &= \mu_j, \quad \forall j \in n, \\ \gamma_{ij} &\geq 0, \quad \forall i, j \in n. \end{aligned}$$

The WD, also known as the Earth Mover's Distance (EMD), is a metric used to quantify the dissimilarity between two probability distributions. Unfortunately, computing WD between two probability distributions over $\mathcal{Y}$ exactly is generally intractable, as it requires an enumeration over $\mathcal{Y}$. Still, WD can be approximated by using empirical distributions with a finite number of samples with the convergence rate of $\mathcal{O}(n^{-\frac{1}{d}})$ (Peyré and Cuturi, 2020).

C Proof of Lemma 1

We start by analyzing the $u_m(y^*) - u_m(\hat{y}^m)$.

We decompose it as follows:

$$\begin{aligned} & u_m(y^*) - \widehat{u}_m(y^*) + \widehat{u}_m(\hat{y}^m) - u_m(\hat{y}^m) + \underbrace{\widehat{u}_m(y^*) - \widehat{u}_m(\hat{y}^m)}_{\leq 0} \\ & \leq \Delta(u_m, \widehat{u}_m, y^*) + \Delta(\widehat{u}_m, u_m, \hat{y}^m) \end{aligned}$$

We can express $\Delta(u_m, \widehat{u}_m, y^*)$ using the following formulation, derived from Lemma 4.

$$\Pr\left(|\Delta(u_m, \widehat{u}_m, y^*)| \leq \epsilon\right) \leq 1 - 2\exp\left(-\frac{2n\epsilon^2}{U_{\max}^2}\right) = 1 - \frac{\delta}{2}.$$

$\Delta(u_m, \widehat{u}_m, y^*)$ holds the following inequality with probability at least $1 - \frac{\delta}{2}$.

$$\Delta(u_m, \widehat{u}_m, y^*) \leq U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{4}{\delta}\right)}.$$

Next, we analyze $\Delta(\widehat{u}_m, u_m, \hat{y}^m)$, however, Lemma 4 cannot be directly applied because $\hat{y}^m$ depends on $\widehat{u}_m$. To address this dependency, we instead utilize Lemma 5, and we can get the following formulation.

$$\Pr\left(\max_y |\Delta(\widehat{u}_m, u_m, y)| \leq 2\mathcal{R}_n(\mathcal{F}) + \epsilon\right) \leq 1 - \exp\left(-\frac{2n\epsilon^2}{U_{\max}^2}\right) = 1 - \frac{\delta}{2}.$$

We can thus express $\Delta(\widehat{u}_m, u_m, \hat{y}^m)$ with probability at least $1 - \frac{\delta}{2}$.

$$\Delta(\widehat{u}_m, u_m, \hat{y}^m) \leq \max_y |\Delta(\widehat{u}_m, u_m, y)| \leq 2\mathcal{R}_n(\mathcal{F}) + U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{2}{\delta}\right)}.$$

From Section 27.2 (Shalev-Shwartz and Ben-David, 2014), the following upper bound on the Rademacher complexity $\mathcal{R}_n(\mathcal{F})$ is obtained under Assumption 1:

$$2\mathcal{R}_n(\mathcal{F}) \leq \frac{12U_{\max}}{n}\left(\sqrt{d\log(2\sqrt{d})} + 2\sqrt{d}\right)$$

From above inequality, we can get the following the bound:

$$\begin{aligned}\Delta(\widehat{u}_m, u_m, \hat{y}^m) &\leq \max_y |\Delta(\widehat{u}_m, u_m, y)| \\ &\leq \frac{12U_{\max}}{n}\left(\sqrt{d\log(2\sqrt{d})} + 2\sqrt{d}\right) + U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{2}{\delta}\right)} \\ &\leq \frac{12U_{\max}}{n}\left(\sqrt{d\log(2\sqrt{d})} + 2\sqrt{d}\right) + U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{4}{\delta}\right)}.\end{aligned}$$

If $d \geq 4$,

$$\begin{aligned}\Delta(\widehat{u}_m, u_m, \hat{y}^m) &\leq \frac{36U_{\max}}{n}\sqrt{d\log(\sqrt{d})} + U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{4}{\delta}\right)}. \\ &\leq 3\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d}.\end{aligned}$$

Otherwise, if $d < 4$,

$$\begin{aligned}\Delta(\widehat{u}_m, u_m, \hat{y}^m) &\leq \frac{36U_{\max}}{n}\sqrt{d\log(\sqrt{d})} + U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{4}{\delta}\right)}. \\ &\leq 3\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{72\sqrt{d}}{n}.\end{aligned}$$

D The Case of P_{human} and P_{model} are identical.

If P_{human} and P_{model} are equal, the upper bound of Regret_n corresponds to Lemma 1:

$$\begin{aligned}\text{Regret}_n &\leq 2U_{\max}\sqrt{\frac{1}{2n}\log\frac{8}{\delta}} + \frac{12U_{\max}}{n}\left(\sqrt{d\log(2\sqrt{d})} + 2\sqrt{d}\right). \\ &\leq 3\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\left(\sqrt{d\log d}\right).\end{aligned}$$

In most studies, the primary goal of MBR decoding studies is to derive y^{m}, given that P_{human} is inaccessible. These studies implicitly assume $P_{\text{model}} = P_{\text{human}}$, highlighting the significance of the results. This finding is for understanding and improving the practical application of MBR decoding methods.

E Proof of Lemma 2

We can derive the following inequality under Assumption 2 for any $y \in \mathcal{Y}$:

$$
\begin{aligned}
\Delta(u_h, u_m, y) &\leq |\Delta(u_h, u_m, y)| \\
&= \left| \sum_{y' \in \mathcal{Y}} P_{\text{human}}(y') u(y, y') - \sum_{y'' \in \mathcal{Y}} P_{\text{model}}(y'') u(y, y'') \right| \\
&= \left| \sum_{y', y''} (u(y, y') - u(y, y'')) \gamma(y', y'') \right| \\
&\text{where} \quad \sum_{y'} \gamma(y', y'') = P_{\text{model}}(y''), \quad \sum_{y''} \gamma(y, y'') = P_{\text{human}}(y'), \quad \gamma(y', y'') \geq 0 \\
&\leq \min_{\gamma} \sum_{y', y''} |u(y, y') - u(y, y'')| \gamma(y', y'') \\
&\leq \min_{\gamma} \sum_{y', y''} C(y', y'') \gamma(y', y'') \\
&= \mathrm{WD}(P_{\text{human}}, P_{\text{model}})
\end{aligned}
$$

F Proof of Lemma 3

Under the Assumption 3, with using the Lemma 4, the $\Delta(u_h, u_m, y^*)$ term is expressed as follows:

$$
\Pr\left(|\Delta(u_h, u_m, y^*)| \leq \epsilon\right) \leq 1 - 2\exp\left(-\frac{2|D|\epsilon^2}{U_{\max}^2}\right) = 1 - \frac{\delta}{2}.
$$

We rearrange ϵ as follows:

$$
\epsilon = U_{\max} \sqrt{\frac{1}{2|D|} \log\left(\frac{4}{\delta}\right)}.
$$

In other words, the upper bound of $\Delta(u_h, u_m, y^*)$ has a probability of at least $1 - \frac{\delta}{2}$.

$$
\Delta(u_h, u_m, y^*) \leq U_{\max} \sqrt{\frac{1}{2|D|} \log\left(\frac{4}{\delta}\right)}.
$$

For the $\Delta(u_h, u_m, \hat{y}^m)$ term, the upper bound can be obtained by the same operation, and the complement Lemma 3 is proved.

$$
\Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, \hat{y}^m) \leq 3 \sqrt{\frac{1}{|D|} \log\frac{1}{\delta}}.
$$

G Regret Bound for MBR decoding with temperature sampling

We have been considering random sampling so far, but we also analyze what the bounds would be if we did temperature sampling, considering practical aspects.

$$
P_{\text{model}}^t(y) = \frac{\exp\left(t^{-1} P_{\text{model}}(y)\right)}{\sum_{y' \in \mathcal{Y}} \exp\left(t^{-1} P_{\text{model}}(y')\right)}
$$

where $t \in \mathbb{R}^+$. The objective equation for MBR decoding of the Monte Carlo estimates using a collection of reference hypotheses $\mathcal{Y}_{\text{ref}}^{\text{n}}$ sampled from the model P_{model}^t is as follows:

$$\widehat{u}_m^t(y) = \frac{1}{|\mathcal{Y}_{\text{ref}}^{\text{n}}|} \sum_{y' \in \mathcal{Y}_{\text{ref}}^{\text{n}}} u(y, y').$$

$$\hat{y}_t^m = \arg\max_{y \in \mathcal{Y}_{\text{ref}}^{\text{n}}} \widehat{u}_m^t(y).$$

Our new objective formulation is defined as:

$$\text{Regret}_{n,D}^t \coloneqq u_h(y^*) - u_h(\hat{y}_t^m). \tag{19}$$

We can derive the upper bound of $\text{Regret}_{n,D}^t$ as follows under the condition Theorem 2 for any $t \in \mathbb{R}^+$.

Corollary 3 (Regret Bound for MBR decoding with temperature sampling)**.** *Under Assumption 1, Assumption 2, Assumption 3 and assuming $d \geq 4$, the regret upper bound of the MBR decoding holds for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$\text{Regret}_{n,D}^t \leq 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d} + \text{WD}(P_{\text{model}}, P_{\text{model}}^t).$$

The case of the little samples n might lead to better performance with using the temperature sampling rather than using P_{model} if t is large. However, the above bound has an extra term WD when performing temperature sampling compared to Theorem 2, so the upper bound of $\text{Regret}_{n,D}^t$ might be improved tighter than Corollary 3's derived from. We discuss this later in this section.

Proof. From the definition of $\text{Regret}_{n,D}^t$:

$$\text{Regret}_{n,D}^t \coloneqq u_h(y^*) - u_h(\hat{y}_t^m).$$

The objective equation for MBR decoding using temperature model distribution can be redefined as:

$$u_m^t(y) = \sum_{y' \in \mathcal{Y}} u(y, y') \cdot P_{\text{model}}^t(y').$$

$$y_t^{\text{m}} = \arg\max_{y \in \mathcal{H}} u_m(y).$$

We can decompose the following terms:

$$\begin{aligned}
\text{Regret}_{n,D}^t &\leq \Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, \hat{y}_t^m) + u_m(y^*) - u_m^t(y^*) + u_m^t(y^*) - \widehat{u}_m^t(y^*) \\
&\quad + \widehat{u}_m^t(\hat{y}_t^m) - u_m^t(\hat{y}_t^m) + u_m^t(\hat{y}_t^m) - u_m(\hat{y}_t^m) \\
&= \Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, \hat{y}_t^m) + \Delta(u_m, u_m^t, y^*) + \Delta(u_m^t, \widehat{u}_m^t, y^*) \\
&\quad + \Delta(\widehat{u}_m^t, u_m^t, \hat{y}_t^m) + \Delta(u_m^t, u_m, \hat{y}_t^m)
\end{aligned}$$

First, the involving the terms u_h, u_m is immediately bounded by Lemma 3 with probability at least $1 - \frac{\delta}{2}$ under Assumption 3.

$$\begin{aligned}
\Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, \hat{y}_t^m) &\leq 2U_{\max}\sqrt{\frac{1}{2|D|}\log\left(\frac{8}{\delta}\right)} \\
&\leq 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}.
\end{aligned}$$

Next we derive $\Delta(u_m, u_m^t, y^*) + \Delta(u_m^t, \widehat{u}_m^t, y^*)$'s upper bound. Under Assummption 2, we can get the following the bound.

$$\Delta(u_m, u_m^t, y^*) + \Delta(u_m^t, u_m, \hat{y}_t^m) \leq 2\mathrm{WD}(P_{\mathrm{model}}, P_{\mathrm{model}}^t)$$

Finally, we derive the upper bound for the terms involving u_m and $\widehat{u}_m^t$. We conduct the sample operation he proof of Lemma 1. We can get the following bound with probability at least $1 - \frac{\delta}{2}$ under Assumption 1 and assuming $d \geq 4$:

$$\begin{aligned}\Delta(u_m^t, \widehat{u}_m^t, y^*) + \Delta(\widehat{u}_m^t, u_m^t, \hat{y}_t^m) &\leq 2U_{\max}\sqrt{\frac{1}{2n}\log\left(\frac{8}{\delta}\right)} + \frac{12U_{\max}}{n}\left(\sqrt{d\log(2\sqrt{d})} + 2\sqrt{d}\right) \\ &\leq 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d}.\end{aligned}$$

□

The upper bound of $\mathrm{Regret}_{n,D}^t$ might be improved We focus on $\Delta(u_m, u_m^t, y^*)$. In this paper, we can derive the upper bound with Wasserstein Distance. However, if P_{model} capture P_{human}, $u_m(y^*) < u_h(y^*)$, but we consider P_{model}^t, it is possible to be $y_m^t = y^*$ with little samples, so $\Delta(u_m, u_m^t, y^*)$ can be negative value. In summary, rather than simply deriving an upper bound on the Wasserstein Distance, this bound could be improved by taking into account a more detailed analysis of the temperature sampling.

H Proof of the Corollary 1

We drive the expected upper bound from high probability upper bound. If we have the regret value R with probability at least $1 - \delta$, we can get the expected upper bound the following the equation with worst-case value U (e.g. when considering the MBR decoding in this paper, worst-case value can be 1.)

$$\textbf{Expected Upper Bound for Regret} = (1 - \delta) \cdot R + \delta \cdot U$$

By applying the above equation to Theorem 1 and Theorem 2, the following upper bound is derived.

$$\mathbb{E}\left[\mathrm{Regret}_n\right] \leq 3\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d} + 2\mathrm{WD}(P_{\mathrm{human}}, P_{\mathrm{model}}) + \delta$$

$$\mathbb{E}\left[\mathrm{Regret}_{n,D}\right] \leq 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d} + 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + \delta$$

I Proof of Corollary 2

Before the proof, we derive the upper bound of the utility function difference.

The expectation difference is:

$$\begin{aligned}&\mathbb{E}\left[u(y, y')\right] - \mathbb{E}\left[u'(y, y')\right] \\ &= \mathbb{E}\left[\boldsymbol{\alpha}(y)^\top \boldsymbol{v}(y')\right] - \mathbb{E}\left[\boldsymbol{\alpha}'(y)^\top \boldsymbol{v}(y')\right] \\ &= \mathbb{E}\left[\left(\boldsymbol{\alpha}(y) - \boldsymbol{\alpha}'(y)\right)^\top \boldsymbol{v}(y')\right].\end{aligned}$$

By applying the Cauchy–Schwarz inequality, we obtain:

$$\begin{aligned}&\mathbb{E}[u(y^*, y')] - \mathbb{E}[u'(y^*, y')] \\ &\leq \|\boldsymbol{\alpha}(y^*) - \boldsymbol{\alpha}'(y^*)\| \cdot \|\mathbb{E}[\boldsymbol{v}(y')]\| \\ &\leq \|\boldsymbol{\alpha}(y^*) - \boldsymbol{\alpha}'(y^*)\|.\end{aligned}$$

Next, we prove the corollary 2.

$$u_h(y^*) - u_h(y') = \Delta(u_h, u_m, y^*) + u_m(y^*) - u_m(y') + \Delta(u_m, u_h, y').$$

From Appendix F, we can get the following bound with probability at least $1 - \frac{\delta}{2}$:

$$\Delta(u_h, u_m, y^*) + \Delta(u_m, u_h, y') \leq 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}.$$

The next step is to prove the remaining conditions.

$$\begin{aligned}
u_m(y^*) - u_m(y') &= \Delta(u_m, u', y^*) + \Delta(u', u_m, y') - u'(y') + u'(y^*) \\
&\leq \Delta(u_m, u', y^*) + \Delta(u', u_m, y') \\
&= \Delta(u_m, \widehat{u}_m, y^*) + \Delta(\widehat{u}_m, u', y^*) + \Delta(\widehat{u}_m, u_m, y') + \Delta(u', \widehat{u}_m, y') \\
&\leq 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \|\boldsymbol{\alpha}(y^*) - \boldsymbol{\alpha}'(y^*)\| + \|\boldsymbol{\alpha}(y') - \boldsymbol{\alpha}'(y')\|.
\end{aligned}$$

Finally, we can get the following the bound under Assumption 1 and Assumption 3 with probability at least $1 - \delta$:

$$\mathrm{Regret}^u_{n,D} \leq 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + \|\boldsymbol{\alpha}(y^*) - \boldsymbol{\alpha}'(y^*)\| + \|\boldsymbol{\alpha}(y') - \boldsymbol{\alpha}'(y')\|.$$

J MAP Decoding Upper Bound

In MAP decoding, our objective is to analyze the difference $P_{\text{human}}(y^*_{\text{MAP}}) - P_{\text{human}}(\hat{y}^m_{\text{MAP}})$. To obtain an upper bound, we decompose $P_{\text{human}}(y^*_{\text{MAP}}) - P_{\text{human}}(\hat{y}^m_{\text{MAP}})$ as follows.

$$\begin{aligned}
P_{\text{human}}(y^*_{\text{MAP}}) - P_{\text{human}}(\hat{y}^m_{\text{MAP}}) = P_{\text{human}}(y^*_{\text{MAP}}) - P_{\text{model}}(y^*_{\text{MAP}}) + P_{\text{model}}(y^*_{\text{MAP}}) - P_{\text{model}}(\hat{y}^m_{\text{MAP}}) \\
+ P_{\text{model}}(\hat{y}^m_{\text{MAP}}) - P_{\text{human}}(\hat{y}^m_{\text{MAP}}).
\end{aligned}$$

We solve the upper bound with Lemma 6, so we bound the difference of distributions with the difference of empirical distributions. We also denote y^- as the value immediately before y.

The following equation holds for all y with probability at least $1 - \frac{2}{\delta}$:

$$\begin{aligned}
|\widehat{P}(y) - P_{\text{model}}(y)| &= \left|\left(\widehat{F}(y) - \widehat{F}\left(y^-\right)\right) - \left(F_{\text{model}}(y) - F_{\text{model}}\left(y^-\right)\right)\right|. \\
&\leq \left(|\widehat{F}(y) - F_{\text{model}}(y)| + \left|\widehat{F}\left(y^-\right) - F_{\text{model}}\left(y^-\right)\right|\right). \\
\max_y |\widehat{P}(y) - P_{\text{model}}(y)| &\leq \max_y \left(|\widehat{F}(y) - F_{\text{model}}(y)| + \left|\widehat{F}\left(y^-\right) - F_{\text{model}}\left(y^-\right)\right|\right) \leq 2\epsilon_1.
\end{aligned}$$

We apply Lemma 6 to the above formulation:

$$\Pr\left(\max_{y\in\mathcal{Y}}\left|\widehat{F}(y) - F_{\text{model}}(y)\right| > \epsilon_1\right) \leq 2\exp\left(-2n\epsilon_1^2\right).$$

Finally, we get the bound with probability at least $1 - \frac{\delta}{2}$:

$$\max_{y\in\mathcal{Y}} |\widehat{P}(y) - P_{\text{model}}(y)| \leq 2\sqrt{\frac{1}{2n}\log\frac{8}{\delta}}.$$

The following inequality holds for $\hat{y}^m_{\text{MAP}}$:

$$P_{\text{model}}(\hat{y}^m_{\text{MAP}}) \geq \widehat{P}(\hat{y}^m_{\text{MAP}}) - 2\epsilon_1.$$

This also applies to y^*_{MAP}:

$$P_{\mathrm{model}}\left(y^*_{\mathrm{MAP}}\right) \leq \widehat{P}\left(y^*_{\mathrm{MAP}}\right) + 2\epsilon_1.$$

From the definition, it is clear that $\widehat{P}(\hat{y}^m_{\mathrm{MAP}}) \geq \widehat{P}\left(y^*_{\mathrm{MAP}}\right)$.

$$\begin{aligned} P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) &\geq \widehat{P}(\hat{y}^m_{\mathrm{MAP}}) - 2\epsilon_1 \\ &\geq \widehat{P}\left(y^*_{\mathrm{MAP}}\right) - 2\epsilon_1 \\ &\geq P_{\mathrm{model}}\left(y^*_{\mathrm{MAP}}\right) - 4\epsilon_1. \end{aligned}$$

Therefore, we can get the upper bound at least $1 - \frac{\delta}{2}$:

$$P_{\mathrm{model}}\left(y^*_{\mathrm{MAP}}\right) - P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) \leq 4\sqrt{\frac{1}{2n}\log\frac{8}{\delta}}.$$

$$\Pr\left(\max_{y\in\mathcal{Y}} |F_{\mathrm{model}}(h) - F_{\mathrm{human}}(h)| > \epsilon_2\right) \leq 2\exp\left(-2|D|\epsilon_2^2\right).$$

We also use Lemma 6. It satisfies with probability at least $1 - \frac{\delta}{2}$:

$$P_{\mathrm{human}}(y^*_{\mathrm{MAP}}) - P_{\mathrm{model}}(y^*_{\mathrm{MAP}}) + P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}) \leq 4\sqrt{\frac{1}{2|D|}\log\frac{8}{\delta}}.$$

Finally, we get the following upper bound with probability at least $1 - \delta$:

$$\begin{aligned} \mathrm{Regret}^{\mathrm{MAP}}_{n,D} &\leq 4\sqrt{\frac{1}{2n}\log\frac{8}{\delta}} + 4\sqrt{\frac{1}{2|D|}\log\frac{8}{\delta}} \\ &\leq 8\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 8\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}. \end{aligned}$$

K Observation

We describe the derivations of the Observation 1 and 2.

K.1 Proof of Observation 1

Assuming the MBR decoding goal is the true value, we aim to know $P_{\mathrm{human}}(h^*) - P_{\mathrm{human}}(\hat{y}^m)$, where $\hat{y}^m$ is the optimal probability based on the empirical distribution of P_{model}, $u_h(h) = \sum P_{\mathrm{human}}(y)u(h,y)$. Remind of $u_h(y^*) - u_h(\hat{y}^m) \leq 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d} = \sigma_1$

$$P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(y^*) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m) \leq P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}) \cdot \sigma_1.$$

Remind of $P_{\mathrm{human}}(h^*) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}) \leq 4\left(\sqrt{\frac{1}{n}} + \sqrt{\frac{1}{|D|}}\right)\left(\sqrt{\frac{1}{2}\log\frac{8}{\delta}}\right)$.

$$P_{\mathrm{human}}(h^*)u_h(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m_{\mathrm{MAP}}) \leq \underbrace{8\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 8\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}}_{\sigma_2}.$$

Combined above formulation:

$$\begin{aligned} &P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(y^*) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m) + P_{\mathrm{human}}(h^*)u_h(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m_{\mathrm{MAP}}) \\ &\quad\leq P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})\sigma_1 + \sigma_2, \\ &P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(y^*) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m_{\mathrm{MAP}}) \\ &\quad\leq P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m) - P_{\mathrm{human}}(h^*)u_h(\hat{y}^m_{\mathrm{MAP}}) + P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})\sigma_1 + \sigma_2, \end{aligned}$$

$$
\begin{aligned}
u_h(y^*) - u_h(\hat{y}^m_{\mathrm{MAP}}) &\leq u_h(\hat{y}^m) - u_h(\hat{y}^m_{\mathrm{MAP}}) + \sigma_1 + \frac{\sigma_2}{P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})}, \\
&\leq u_h(\hat{y}^m) - u_m(\hat{y}^m) + u_m(\hat{y}^m) - u_m(\hat{y}^m_{\mathrm{MAP}}) + u_m(\hat{y}^m_{\mathrm{MAP}}) - u_h(\hat{y}^m_{\mathrm{MAP}}) \\
&+ \sigma_1 + \frac{\sigma_2}{P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})}, \\
&\leq u_m(\hat{y}^m) - u_m(\hat{y}^m_{\mathrm{MAP}}) + 2\sqrt{\frac{1}{2|D|}\log\frac{8}{\delta}} + \sigma_1 + \frac{\sigma_2}{P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})}, \\
&\leq u_m(\hat{y}^m) - u_m(\hat{y}^m_{\mathrm{MAP}}) + O\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{D}}\right)\right).
\end{aligned}
$$

From this, we assume that the MAP decoding target is the true value.

$$
P_{\mathrm{human}}(\hat{y}^m)u_h(y^*) - P_{\mathrm{human}}(\hat{y}^m)u_h(\hat{y}^m) \leq P_{\mathrm{human}}(\hat{y}^m)\cdot\sigma_1.
$$

$$
P_{\mathrm{human}}(y^*_{\mathrm{MAP}})u_h(\hat{y}^m) - P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m) \leq u_h(\hat{y}^m)\cdot\sigma_2.
$$

Combined above formulation:

$$
\begin{aligned}
P_{\mathrm{human}}(y^*_{\mathrm{MAP}})u_h(\hat{y}^m) - P_{\mathrm{human}}(\hat{y}^m)u_h(\hat{y}^m) \leq\ & P_{\mathrm{human}}(\hat{y}^m)\sigma_1 + u_h(\hat{y}^m)\sigma_2 + P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}})u_h(\hat{y}^m) \\
& - P_{\mathrm{human}}(\hat{y}^m)u_h(y^*)
\end{aligned}
$$

$$
\begin{aligned}
P_{\mathrm{human}}(y^*_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m) &\leq P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{human}}(\hat{y}^m)\frac{u_h(y^*)}{u_h(\hat{y}^m)} + \frac{P_{\mathrm{human}}(\hat{y}^m)}{u_h(\hat{y}^m)}\sigma_1 + \sigma_2 \\
&\leq P_{\mathrm{human}}(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) + P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{model}}(\hat{y}^m) \\
&+ P_{\mathrm{model}}(\hat{y}^m) - P_{\mathrm{human}}(\hat{y}^m) + \frac{\sigma_1}{u_h(\hat{y}^m)} + \sigma_2 \\
&\leq 4\sqrt{\frac{1}{2|D|}\log\frac{8}{\delta}} + \frac{\sigma_1}{u_h(\hat{y}^m)} + \sigma_2 + P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{model}}(\hat{y}^m) \\
&\leq O\left(\max\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{D}}\right)\right) + P_{\mathrm{model}}(\hat{y}^m_{\mathrm{MAP}}) - P_{\mathrm{model}}(\hat{y}^m).
\end{aligned}
$$

K.2 Proof of Observation 2

Remid of $\mathrm{Regret}^{\mathrm{MAP}}_{n,D}$ and $\mathrm{Regret}_{n,D}$'s upper bound:

$$
\mathrm{Regret}^{\mathrm{MAP}}_{n,D} \leq \underbrace{8\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 8\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}}}_{\phi_1}.
$$

$$
\mathrm{Regret}_{n,D} \leq \underbrace{4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} + \frac{36}{n}\sqrt{d\log d}}_{\phi_2}.
$$

$$
\phi_1 - \phi_2 = 4\sqrt{\frac{1}{n}\log\frac{1}{\delta}} + 4\sqrt{\frac{1}{|D|}\log\frac{1}{\delta}} - \frac{36}{n}\sqrt{d\log d}.
$$

From the above inequality, we can derive this observation.

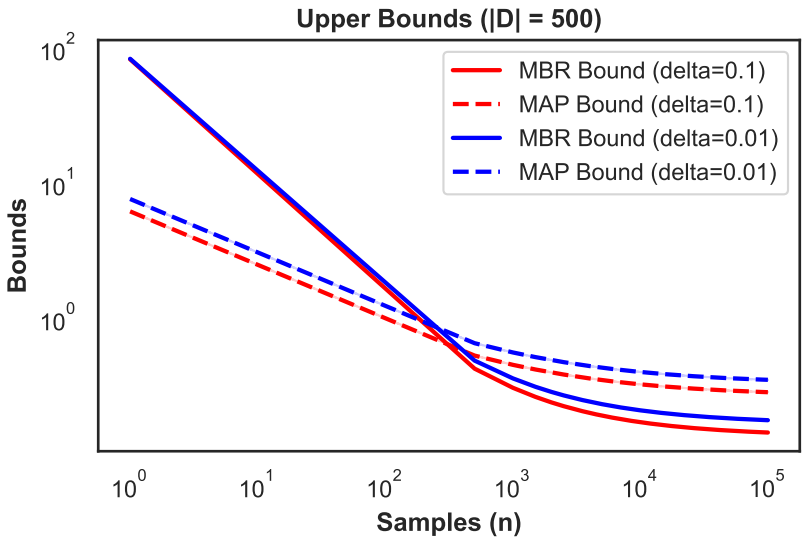

Figure 5: Numerical Experiment for Observation 2's 3

- $n \to \infty$ and D is finite, $\text{Regret}_{n,D} < \text{Regret}_{n,D}^{\text{MAP}}$.
- D is infinite, n is finite, $\text{Regret}_{n,D} < \text{Regret}_{n,D}^{\text{MAP}}$.

$$\frac{1}{9}\sqrt{n \log \frac{1}{\delta}} \geq \sqrt{d \log d}.$$

- D is finite and n is finite, the upper bound of $\text{Regret}_{n,D}$ remains less than or equal to the upper bound of $\text{Regret}_{n,D}^{\text{MAP}}$, provided the following condition holds:

$$\frac{n}{9}\left(\sqrt{\frac{1}{n} \log \frac{1}{\delta}} + \sqrt{\frac{1}{|D|} \log \frac{1}{\delta}}\right) \geq \sqrt{d \log d}.$$

L Observation 2 Experiment

This result shows that the upper bound of MBR decoding is less than the upper bound of MAP decoding at the number of samples described in Observation 2's 3).

M Experimental Details of Numerical Simulation (Section 6)

P_{human} is non-uniform distribution (reflecting real-world biases), for each seed, we generate P_{human} via i.i.d. sampling, then form the empirical model distribution P_{model} by drawing D times from P_{human} (i.e, $\widehat{P}$ represents hypothesis frequencies). In the experiment setting of Fig. 2, we applied the utility function according to Assumption 1, and in Fig. 3 and Fig. 4, we assume a symmetric utility matrix $u \in \mathbb{R}^{\mathcal{Y} \times \mathcal{Y}}$ with $u(i,i) = 1$ and $u(i,j) \in [0,1]$ for $i \neq j$, assigning slightly higher utilities to outcomes with higher P_{human} probabilities.