Revisiting Self-Consistency from Dynamic Distributional Alignment Perspective on Answer Aggregation

Yiwei Li[1*], **Ji Zhang**[1*], **Shaoxiong Feng**[2], **Peiwen Yuan**[1], **Xinglin Wang**[1],
Jiayi Shi[1], **Yueqi Zhang**[1], **Chuyi Tan**[1], **Boyuan Pan**[2], **Yao Hu**[2], **Kan Li**[1†]
[1] School of Computer Science, Beijing Institute of Technology
[2] Xiaohongshu Inc
{liyiwei,jizhang,peiwenyuan,wangxinglin}@bit.edu.cn
{shaoxiongfeng2023}@gmail.com {panboyuan,xiahou}@xiaohongshu.com
{shijiayi,zhangyq,tanchuyi,likan}@bit.edu.cn

Abstract

Self-consistency improves reasoning by aggregating diverse stochastic samples, yet the dynamics behind its efficacy remain underexplored. We reframe self-consistency as a dynamic distributional alignment problem, revealing that decoding temperature not only governs sampling randomness but also actively shapes the latent answer distribution. Given that high temperatures require prohibitively large sample sizes to stabilize, while low temperatures risk amplifying biases, we propose a confidence-driven mechanism that dynamically calibrates temperature: sharpening the sampling distribution under uncertainty to align with high-probability modes, and promoting exploration when confidence is high. Experiments on mathematical reasoning tasks show this approach outperforms fixed-diversity baselines under limited samples, improving both average and best-case performance across varying initial temperatures without additional data or modules. This establishes self-consistency as a synchronization challenge between sampling dynamics and evolving answer distributions.

1 Introduction

Self-consistency (Wei et al., 2022) is a well-established decoding method that enhances model performance by aggregating multiple stochastic samples via majority voting. It has been demonstrated to be highly effective across a variety of tasks (Chen et al., 2023; Wang et al., 2024b), particularly in improving reasoning abilities (Wei et al., 2022). Despite its empirical success, the underlying mechanisms behind self-consistency remain underexplored. In this work, we revisit self-consistency from a distributional perspective, reframing it as a dynamic alignment problem, to achieve more robust and effective performance in answer aggregation.

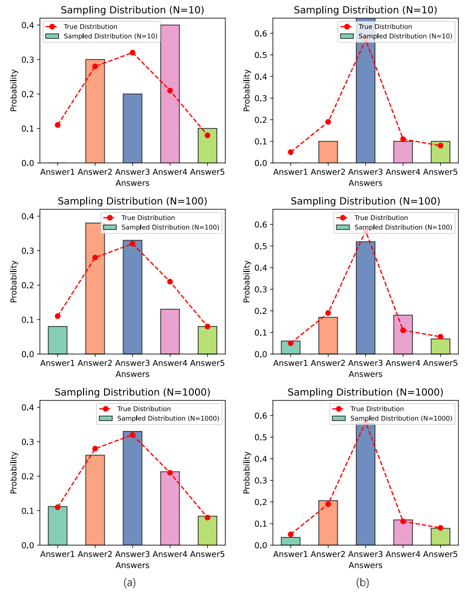


Figure 1: (a) Multiple stochastic sampling for fitting the true distribution. As the sample size increases, the noise gradually diminishes, and ultimately, the top-1 sampled answer aligns with the true distribution. (b) As the temperature decreases, the confidence in the true distribution increases, allowing alignment with the true distribution to be achieved with fewer samples.

Recent work (Wu et al., 2024; Li et al., 2024c) argue that by combining different reasoning traces via majority voting, self-consistency can avoid local optima and reduce the high variance associated with single-sample outputs, ultimately converging to the model's true answer distribution (see Figure 1 (a)). Building on this insight, our work provides a formal definition of its convergence and derives practical criteria for the assessment. Through our convergence analysis, we reveal that this conventional view is limited to a fixed true distribution, overlooking the crucial impact that parameter-controlled decoding (typically the temperature) has on the true distribution (see Figure 1 (b)). Moreover, practical applications are often constrained

[*] Equal contribution.
[†] Corresponding author.

Findings of the Association for Computational Linguistics: ACL 2025, pages 25208–25223
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

by the sampling. Therefore, we raise two key questions: (1) Alignment under Constraints: How does decoding diversity affect the alignment between the sampling distribution and the true answer distribution when only a limited number of samples is available? (2) Dynamic Alignment: Can we actively calibrate the diversity in practice to accelerate and stabilize convergence, rather than passively waiting for asymptotic convergence?

To explore these issues, we analyze the impact of diversity on self-consistency. The temperature parameter not only governs the randomness of sampling but also directly shapes the true answer distributions. Our findings reveal that as the number of samples approaches infinity, a higher temperature yields a more ideal true answer distribution. However, when the sample size is finite, the optimal sampling temperature decreases as the number of samples diminishes. This leads to a trade-off: low-diversity sampling quickly concentrates the answers and suppresses noise but risks amplifying model biases, whereas high-diversity sampling disperses the answers, requiring more samples to stabilize, yet it enables the exploration of a potentially superior true distribution.

In summary, our comprehensive analysis indicates that the effectiveness of self-consistency hinges on a dynamic alignment between the confidence of the sampling distribution and the intrinsic uncertainty of the true answer distribution—a relationship that is influenced by the number of samples. Ideally, the sampling distribution should be controlled such that the majority voting outcomes closely match the true distribution, and on this basis, explores toward an improved true distribution.

Based on this insight, we propose a confidence-driven diversity optimization mechanism that dynamically adjusts the temperature based on real-time confidence values derived from the answer distribution. When early samples show only a small probability gap between the top two most-voted answers, our mechanism sharps the sampling distribution to better align it with the true distribution. Conversely, when confidence is high, the temperature is increased to explore potentially superior distributions. We derive a confidence threshold to determine the direction of temperature adjustment, providing theoretical support for this process. This closed-loop control dynamically synchronizes the sampling distribution with the latent answer distribution, ensuring efficient convergence while actively pursuing a better distribution. Experimental results across various model types and size indicate that it can achieve simultaneous improvements in both average and best performance across different initial temperatures, without requiring any additional training, valid data, reward models, or external modules.

2 Fundamental Analysis of Self-Consistency

In this section, we first present a distribution alignment perspective on how self-consistency works with specific true answer distributions, supported by experimental evidence to substantiate this viewpoint. Building upon this foundation, we proceed to provide both a formal definition of self-consistency convergence and practical criteria for assessment.

2.1 Why Self-Consistency Works: A Distributional Perspective

Self-Consistency is a widely-used decoding method for improving reasoning performance by aggregating multiple stochastic samples. By applying a majority voting scheme, it mitigates issues such as local optima and high variance that arise from relying on a single sample. Formally, it can be expressed as:

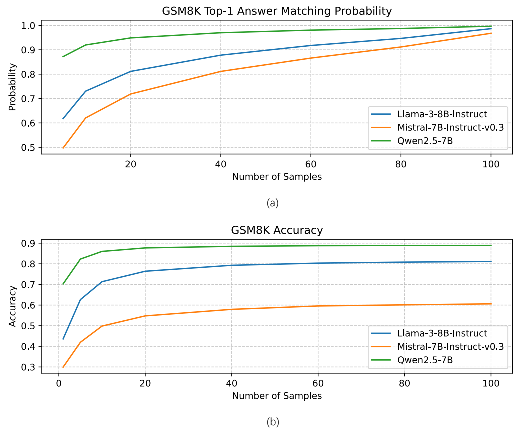
$$\hat{y}_{SC} = \arg\max_{y} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(y_i = y) \right) \tag{1}$$

where y_i is the i-th sampled answer, and $\mathbb{I}(y_i = y)$ is the indicator function that equals 1 if y_i matches the candidate answer y, and 0 otherwise. The result, $\hat{y}_{SC}$, is the answer with the highest number of votes (the top-1 answer).

From a probabilistic perspective, self-consistency can be seen as a *Monte Carlo estimator* of the true answer distribution $p(y \mid \mathbf{x})$. As the number of samples increases, the empirical distribution formed by the samples approximates the true distribution, and the most frequent answer aligns with the true distribution:

$$\begin{aligned} \hat{p}_{SC}(y) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(y_i = y) \\ &\to p(y \mid \mathbf{x}), \quad \text{as} \quad n \to \infty \end{aligned} \tag{2}$$

As the number of samples increases, the estimation becomes more reliable, and the voting mechanism converges towards the true answer.



(a)

(b)

Figure 2: Top-1 answer matching probability (a) and accuracy (b) both improve as the sampling number increases.

Experimental Analysis To validate this viewpoint, we analyzed the top-1 answer match rate as a function of the sample size. The true top-1 answer is simulated by drawing from a large sample to approximate the true distribution. Results from Figure 2 reveals **Findings 1**: As the sample size increased, the top-1 answer match rate gradually approaches 100% with the accuracy consistently improves. Based the observation, we derive the following insight: ***Insights 1****: The improvement in self-consistency performance stems from the fact that, the top-1 answer in the sampling distribution gradually aligns with the true distribution, ultimately enhancing accuracy to match the true distribution's level.*

2.2 Convergence Analysis of Answer Aggregation

According to ***Insights 1***, since the accuracy of the true distribution is fixed, the performance of self-consistency is guaranteed to converge. To further investigate it, we provide the following definition according to the Cauchy convergence criterion:

Definition 2.1. *Let $f^M(i) = \sum_{l=1}^{M} \mathbb{I}(\hat{y}_l = i)$, where $\hat{y}_l$ represents the set of answers generated by the model, and M is the number of samples. For any given $\epsilon > 0$, there exists a positive integer L such that for $N, M > L$, if the following holds:*

$$\left| \underset{i}{argmax}\ f^M(i) - \underset{i}{argmax}\ f^N(i) \right| < \epsilon \quad (3)$$

we can conclude that self-consistency has converged.

Based on Definition 2.1, we prove that self-consistency also converges in terms of the accuracy

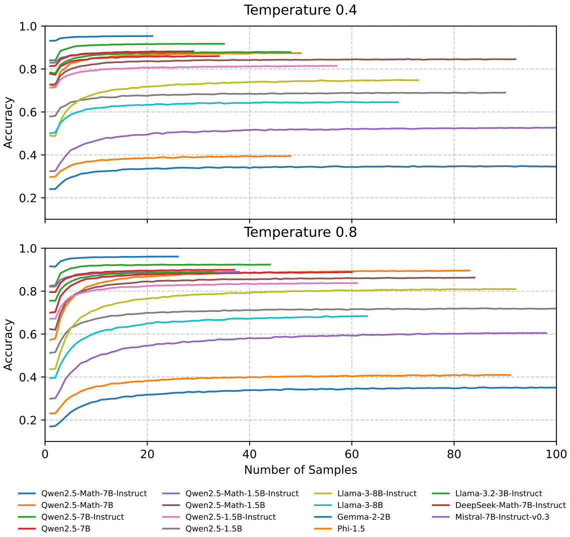


Figure 3: Self-consistency convergence plots under different temperature (0.4 and 0.8) settings.

on the dataset:

Theorem 2.2. *Let $Acc_D^M = \frac{1}{|D|} \sum_{j \in D} \mathbb{I}[\underset{i}{argmax}\ f^M(i) = gt_j]$ denote the accuracy of self-consistency when a single question is sampled M times on dataset D, where gt_j represents the correct answer to the j-th question. If Definition 1 holds, then for any given $\epsilon > 0$, there exists a positive integer L such that when $N, M > L$, the following holds:*

$$\left| Acc_D^M - Acc_D^N \right| < \epsilon \quad (4)$$

The Proof of Theorem 2.2 is in Appendix A. By setting ϵ to $\frac{1}{|D|}$, the following definition is established:

Definition 2.3. *If the following holds on dataset D:*

$$\left| Acc_D^M - Acc_D^{M-5} \right| < \frac{1}{|D|} \quad (5)$$

we can consider self-consistency to have converged at a sample size of M.

Experimental Analysis Figure 3 depicts the convergence behavior of various models, with the accuracy curves plotted up to the convergence point according to Definition 2.3, from where we can get: **Findings 2**: The convergence speed exhibits a positive correlation with accuracy. **Findings 3**: The convergence speed is inversely correlated with temperature. **Findings 4**: The final converged accuracy

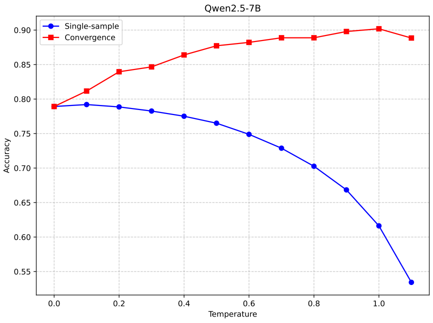


Figure 4: The accuracy curve with varying temperature under convergence.

varies across different temperature settings. Based on them, we derive ***Insights 2***: *Sampling diversity will affect the true distribution, impacting both the convergence accuracy and the convergence speed of self-consistency.*

3 Diversity Trade-offs for Self-Consistency

3.1 Sampling Diversity Affection

According to ***Insight 2***, to gain a deeper understanding of the impact of diversity on self-consistency, we investigate how accuracy varies with temperature changes in increments of 0.1. The study is divided into two parts: convergence analysis and finite-sample analysis.

Converge Condition Figure 4 indicates **Findings 5**: As the temperature increases, the accuracy of single samples exhibits a declining trend, while the accuracy of self-consistency after convergence shows an increasing trend (the optimal point is often near 1.0[1]). Please refer to Appendix B for more results. The disagreement resolution theorem in ensemble learning provides a potential explanation, suggesting that the overall performance of an ensemble is determined by the trade-off between the accuracy of individual models and the diversity among them. From this trend and ***Insights 1***, we gain ***Insights 3***: *When the sample size is sufficient, the temperature should be increased to better explore the true distribution with higher accuracy.*

Finite-Sample Condition Figure 5 indicates **Findings 6**: When the sample size is limited, the optimal temperature gradually shifts toward lower values as the sample size decreases. Please refer to Appendix B for more results. This findings and ***Insights 1*** leads us to ***Insights 4***: *Sample size determines the maximum top-1 confidence level that can be reliably modeled. True distributions with lower confidence require larger data volumes to ensure that the sampled top-1 answer aligns with the converged result.*

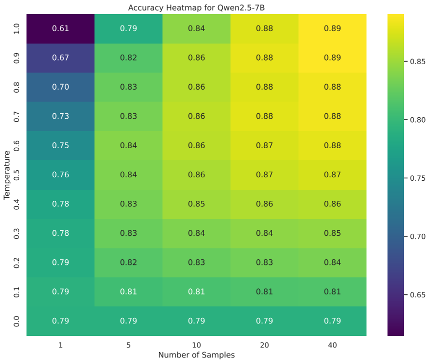


Figure 5: The accuracy heatmap with varying temperature with finite sample size.

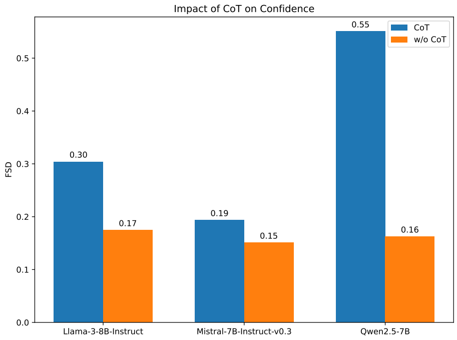


Figure 6: FSD (Equation 7) (Lyu et al., 2024) is employed as the confidence metric to quantify the gap between top two candidates.

By combining ***Insights 3*** and ***4***, we can derive ***Insights 5***: *The effectiveness of self-consistency depends on dynamically aligning the confidence of the sampling distribution with the inherent uncertainty of the true answer distribution.*

3.2 Chain-of-thought Affection

Besides the sampling diversity decided by temperature, Chain-of-Thought (Wei et al., 2022) is also a key factor. From Figure 6 we can get **Findings 7**: Using CoT prompt leads to higher confidence compared to not using it. A deeper ***Insight 6*** emerges: *Chain-of-thought (CoT) reasoning narrows the output space and reduces diversity, thereby increasing answer confidence.* However, investigating this phenomenon is not the focus of this paper, and we

[1] We speculate that this may be related to the training temperature being typically set to 1.0. We leave the study of the optimal temperature as future work.

leave it for future work.

4 From Static to Adaptive: Confidence-Driven Optimization of Self-Consistency Distributions

4.1 Motivation

Traditional self-consistency methods rely on static sampling strategies with fixed sample sizes and temperatures, which face limitations: When the confidence of the sampling distribution is too low, a limited sample size struggles to accurately capture the true top-1 answer, restricting the performance gains of self-consistency (***Insight 4***). Conversely, when the confidence is too high, the model fails to explore potentially better distributions at higher temperatures (***Insight 3***). Our method addresses these issues through adaptive confidence-distribution alignment (***Insight 5***). By dynamically adjusting the sampling distribution's diversity based on real-time confidence levels, we optimize alignment by proactively adapting to the evolving gap between model confidence and true distribution uncertainty. This dynamic mechanism enables efficient convergence to the correct answer even under limited sample sizes while facilitating exploration of better true distributions when needed. Through this approach, we enhance both the accuracy and robustness of self-consistency across diverse conditions.

4.2 Diversity Control Strategy

Dynamic Temperature Adjustment We introduce a confidence-driven diversity optimization mechanism to dynamically align the sampling distribution with the latent answer distribution. First-Second Distance (FSD) (Lyu et al., 2024) is employed as the confidence metric to quantify the gap between top candidates. Formally, at decoding step t:

$$\mathrm{FSD}^{(t)} = {p_1}^{(t)} - {p_2}^{(t)} \tag{6}$$

where ${p_1}^{(t)}$ and ${p_2}^{(t)}$ are the probabilities of the top two answers from the first t samples. This metric directly reflects the model's uncertainty in distinguishing between the dominant candidates.

To ensure stable optimization, we design a conservative adjustment rule with a dead zone around confidence threshold τ. The temperature T is updated based on the FSD metric:

$$T^{(t+1)} = \begin{cases} T^{(t)} - 0.1 & \text{if } \mathrm{FSD}^{(t)} < \tau - \epsilon, \\ T^{(t)} + 0.1 & \text{if } \mathrm{FSD}^{(t)} > \tau + \epsilon, \\ T^{(t)} & \text{otherwise,} \end{cases} \tag{7}$$

where ϵ is a stability margin, which we set to 0.05 for simplicity. Temperature T is clamped to [0.1, 1.0] to avoid extreme values.

Phased Sampling Strategy To balance exploration and efficiency, we employ a three-phase sampling protocol:

- Exploration Phase: Collect small number of samples ($n_1 = 5$) with preset $T^{(1)}$ as a window to estimate initial FSD$^{(1)}$.
- Adaptive Phase: Adjust $T^{(2)}$ through Equation 7, then generate $n_2 = 0.5N - n_1$ (N: total budget) additional samples.
- Exploitation Phase: Finalize $T^{(3)}$ and generate the remaining $n_3 = 0.5N$ samples.

The phased approach progressively shifts from broad exploration to focused exploitation. Finally, the accuracy is calculated by majority voting from the total of N samples.

In summary, our method dynamically adjusts the sampling diversity by monitoring the confidence levels, allowing for more efficient exploration and convergence. This adaptive mechanism ensures better alignment with the true answer distribution under sampling constraints to improve accuracy.

4.3 Theoretical Analysis

To ensure a rational and effective selection of the FSD threshold τ, we construct a one-sided z-test for analysis. The test employs the null hypothesis as follows:

H_0: The current sampled top-1 answer is not the true answer for the given question under infinite sampling.

To simplify this problem, we assume that only the current top-2 answer could potentially become the true answer under infinite sampling. Consequently, it is natural to focus on the relationship between FSD and confidence. Therefore, this hypothesis can be described as:

$$z = \frac{\hat{d} - d_\mu}{SE} \tag{8}$$

where $\hat{d}$ represents the observed FSD from actual sampling, d_μ denotes the FSD under the true distribution of the model, and SE stands for the standard error. Based on the null hypothesis H_0, it is clear that $d_\mu < 0$. We adopt the critical condition $d_\mu = 0$:

$$z \geq \frac{\hat{d} - 0}{SE} = \frac{\hat{p}_1 - \hat{p}_2}{SE} \tag{9}$$

Assuming that the current sample size approaches infinity and that the sampling between the two categories can be considered independent, according to the multinomial distribution and Jensen's inequality (in the case of a concave function), we have:

$$\begin{aligned} SE &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{N} + \frac{\hat{p}_2(1-\hat{p}_2)}{N}} \\ &\leq \sqrt{\frac{2p(1-p)}{N}} \end{aligned} \tag{10}$$

where $p = \frac{\hat{p}_1+\hat{p}_2}{2} \in (0, 0.5]$, substituting Equation 10 into Equation 9, we can derive the theoretical lower bound of z:

$$z \geq \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{2p(1-p)}{N}}} \geq \hat{d}\sqrt{2N} \tag{11}$$

Setting $z = 1.64$, the corresponding p-value is approximately 0.05, indicating strong statistical evidence that the current top-1 answer is indeed the top-1 answer under the true model distribution. Therefore, the FSD threshold can be set as:

$$\tau = \frac{1.16}{\sqrt{N}} \tag{12}$$

5 Experiments

5.1 Experiment Setup

Datasets and Models We evaluate our method on two widely-used mathematical reasoning benchmarks: GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021). Experiments span multiple model families to assess generalizability, including Qwen (Yang et al., 2024), Llama (Dubey et al., 2024), Mistral(Jiang et al., 2023), DeepSeek(Shao et al., 2024), Gemma(Rivière et al., 2024) and Phi(Li et al., 2023b).

Implementation Details To systematically compare dynamic versus static temperature strategies, we test initial temperatures $T_0 \in \{0.1, 0.2, ..., 1.0\}$ with sampling budgets $N \in \{10, 20, 40\}$.

Metric To provide an intuitive and efficient evaluation of the differences between methods, we calculate both the average and maximum accuracy for fixed-temperature sampling and dynamic-temperature sampling across all temperatures. The evaluation is conducted from the perspectives of robustness and effectiveness. Formally:

$$Mean = \frac{1}{N_T} \sum_{t \in T_0} Acc_t \tag{13}$$

$$Max = \max_{t \in T_0} Acc_t \tag{14}$$

5.2 Results

From the results presented in Table 1 through 15 models, we can find:

Dynamic temperature sampling mitigates the performance degradation associated with fixed-temperature sampling. We find that the average accuracy across different temperatures for dynamic temperature sampling outperforms fixed-temperature sampling in the majority of models. This suggests that our method is not constrained by the temperature range and can identify samples that are more effective for self-consistency performance across various temperatures. To some extent, this approach mitigates the performance loss caused by ineffective sampling at a single fixed temperature.

For different samples, dynamic temperature sampling searches for a more suitable temperature for each sample. Similarly, we observe that dynamic temperature sampling also provides a certain improvement in terms of the maximum accuracy. This can be attributed to the fact that different samples require different temperature ranges. Fixed-temperature sampling can only achieve the desired accuracy for the dataset as a whole, whereas dynamic temperature sampling automatically searches for a more optimal temperature for each individual sample, maximizing the performance of self-consistency optimization across various temperatures.

5.3 Analysis

Visualization We provide a detailed analysis of the model's accuracy at different temperatures. Figure 7 presents the accuracy and temperature curve for the Qwen2.5-Math-7B model. We observe that, with sampling sizes of 20 and 40, both low temperature ranges (0.1-0.4) and high temperature ranges (0.7-1.0) exhibit notable improvements. This suggests that dynamic temperature sampling yields

Models	Strategy	GSM8K						MATH					
		N=10		N=20		N=40		N=10		N=20		N=40	
		Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Qwen2.5-1.5B	Fix	65.4	67.6	67.2	69.6	68.2	**70.9**	31.4	**36.1**	34.5	38.6	36.5	40.8
	Dynamic	**65.7**	**67.7**	**67.8**	**69.8**	**68.9**	**70.9**	**32.4**	36.0	**36.5**	**38.9**	**38.7**	**41.0**
Qwen2.5-1.5B-Instruct	Fix	79.0	**80.7**	80.3	82.2	81.1	83.2	51.9	**53.8**	53.3	55.0	54.1	55.8
	Dynamic	**79.2**	**80.7**	**80.8**	**82.4**	**81.6**	**83.5**	**52.3**	53.6	**53.9**	**55.2**	**54.6**	**55.9**
Qwen2.5-7B	Fix	84.6	86.1	85.7	87.7	86.3	88.9	48.7	52.0	50.7	53.8	51.8	54.9
	Dynamic	**84.7**	**86.3**	**86.1**	**88.1**	**86.8**	**89.0**	**49.6**	**52.3**	**51.9**	**53.9**	**53.2**	**55.1**
Qwen2.5-7B-Instruct	Fix	**90.8**	91.9	91.2	92.2	91.4	**92.4**	65.9	66.6	66.6	**67.3**	66.9	**67.7**
	Dynamic	**90.8**	**92.0**	**91.4**	**92.3**	**91.7**	**92.4**	**66.1**	**66.7**	**66.8**	**67.3**	**67.2**	67.6
Qwen2.5-Math-1.5B	Fix	80.1	**83.3**	82.0	84.6	82.9	**85.6**	41.5	44.1	43.1	**46.0**	44.2	47.1
	Dynamic	**80.7**	83.2	**82.9**	**84.7**	**83.9**	**85.6**	**41.9**	**44.2**	**44.1**	**46.0**	**45.2**	**47.2**
Qwen2.5-Math-1.5B-Instruct	Fix	87.1	88.2	87.7	**88.8**	87.9	89.0	64.2	65.1	64.7	**65.8**	64.9	**66.1**
	Dynamic	**87.2**	**88.4**	**87.9**	**88.8**	**88.2**	**89.2**	**64.3**	**65.2**	**64.8**	65.6	**65.1**	65.9
Qwen2.5-Math-7B	Fix	82.2	85.0	84.6	87.1	85.8	88.2	52.7	56.1	54.9	57.9	56.3	59.4
	Dynamic	**83.0**	**85.4**	**85.5**	**87.4**	**86.8**	**88.5**	**53.4**	**56.2**	**56.2**	**58.4**	**57.7**	**59.7**
Qwen2.5-Math-7B-Instruct	Fix	94.9	95.8	95.2	**96.0**	95.4	**96.2**	68.8	70.1	69.5	**70.9**	69.7	**70.9**
	Dynamic	**95.1**	**95.9**	**95.4**	**96.0**	**95.6**	**96.2**	**69.3**	**70.4**	**70.0**	70.7	**70.2**	**70.9**
Llama-3-8B	Fix	58.2	63.0	60.9	65.8	62.5	67.4	18.6	21.5	20.3	23.5	21.7	25.1
	Dynamic	**59.3**	**63.4**	**62.6**	**66.1**	**64.3**	**67.6**	**19.3**	**21.9**	**22.1**	**24.2**	**23.6**	**25.5**
Llama-3-8B-Instruct	Fix	66.6	72.0	70.2	76.1	72.2	78.6	**20.1**	24.4	21.3	26.8	22.1	28.7
	Dynamic	**67.1**	**72.7**	**71.6**	**76.9**	**74.1**	**79.5**	**20.1**	**25.0**	**21.4**	**26.9**	**22.3**	**28.8**
Gemma-2-2B	Fix	29.1	32.2	31.0	33.9	32.3	**34.9**	14.6	16.5	16.1	**18.2**	16.8	18.2
	Dynamic	**29.7**	**32.3**	**32.3**	**34.2**	**33.5**	34.7	**15.1**	**16.7**	**16.8**	17.9	**17.6**	**18.4**
Phi-1.5	Fix	35.1	37.6	37.0	39.5	38.1	40.7	4.0	4.7	4.6	5.0	5.0	5.5
	Dynamic	**35.6**	**37.7**	**37.8**	**39.6**	**39.0**	**40.8**	**4.2**	**5.0**	**4.8**	**5.2**	**5.3**	**5.7**
DeepSeek-Math-7B-Instruct	Fix	**87.4**	**88.6**	88.1	89.5	88.5	**90.1**	44.4	45.8	46.2	**48.2**	47.1	**49.5**
	Dynamic	**87.4**	**88.6**	**88.3**	**89.8**	**88.7**	**90.1**	**44.8**	**46.1**	**46.6**	**48.2**	**47.8**	**49.5**
Llama-3.2-3B-Instruct	Fix	**86.2**	**87.4**	87.2	88.4	87.7	88.9	48.7	49.8	50.2	51.4	51.2	52.4
	Dynamic	**86.2**	87.3	**87.5**	**88.6**	**88.1**	**89.2**	**49.0**	**50.1**	**50.6**	**51.6**	**51.7**	**52.7**
Mistral-7B-Instruct-v0.3	Fix	46.1	48.9	49.3	53.3	51.5	57.2	17.1	18.3	19.2	20.8	20.8	22.4
	Dynamic	**46.6**	**49.7**	**50.4**	**55.0**	**52.8**	**58.9**	**17.6**	**19.0**	**20.2**	**21.0**	**22.2**	**23.5**

Table 1: Evaluation results by using 15 models from different base architectures on GSM8K(Cobbe et al., 2021) and MATH(Hendrycks et al., 2021). Dynamic temperature sampling achieves superior average and maximum performance across a wide range of settings.

more robust results. However, with a sampling size of 10, the performance in the low temperature range is almost identical to that of fixed-temperature sampling, primarily due to the constraints of the sample size. In the more optimal temperature range (0.4-0.7), the performance of dynamic and fixed-temperature sampling is similar, which aligns with our expectations and indicates that the intermediate temperature has already achieved a balanced trade-off.

Direction Analysis of Temperature Variation Taking the sample level into account, we first analyzed the proportions of samples that experienced temperature increases, decreases, or remained constant throughout the dynamic temperature sampling process, as illustrated in Figure 8. We observed that in the low temperature range, at least 80% of the samples experienced an increase in temperature. This observation is consistent with our hypothesis derived from dataset-level considerations, which suggests that increasing the temperature tends to result in higher expected accuracies. As the temperature rises, the proportion of samples experiencing temperature increases gradually declines, indicating that for some samples at the current sampling size, excessive temperatures are insufficient to confidently select the correct answer. Consequently, lowering the temperature becomes necessary to enhance FSD. Additionally, we noticed that with higher sampling sizes, the proportion of samples

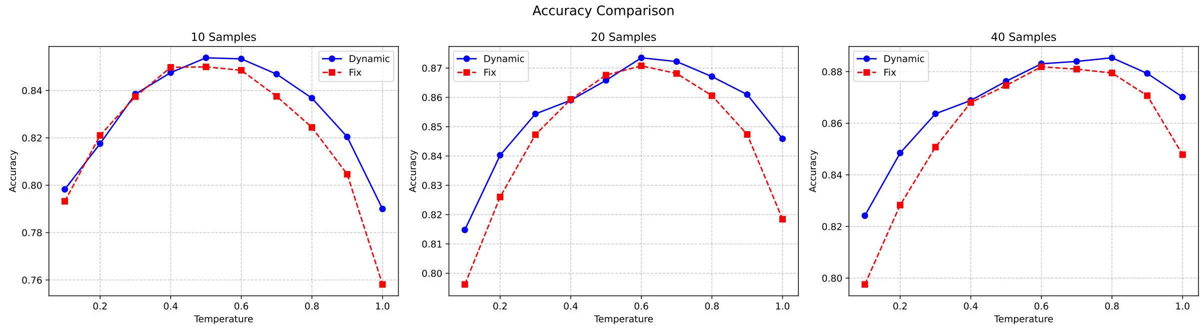


Figure 7: A detailed results of the model's accuracy across different temperatures. Our method achieves better performance under both lower and higher initial temperatures.

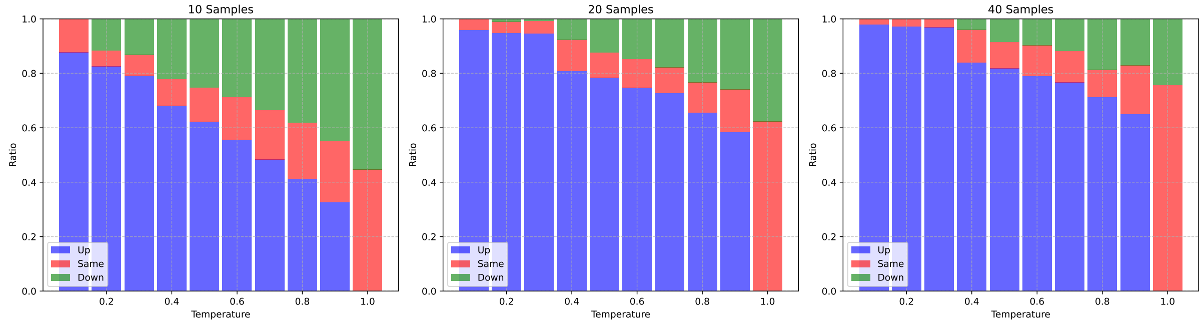


Figure 8: Proportions of samples with temperature increases, decreases, or stability during dynamic temperature sampling.

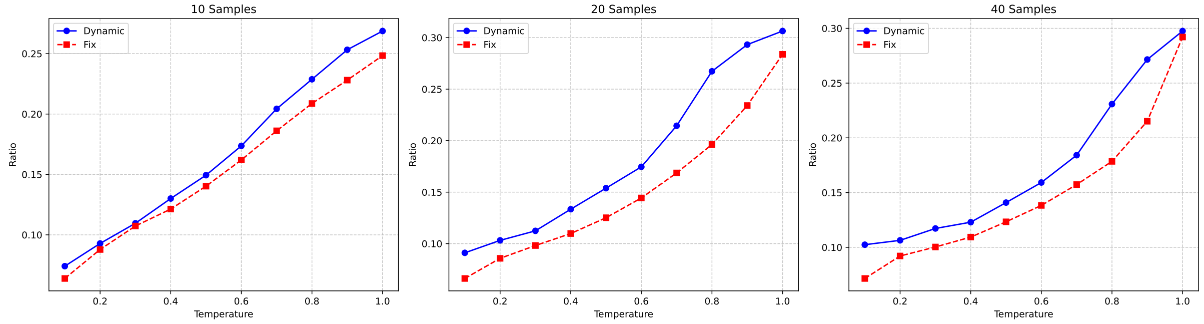


Figure 9: Proportion of FSD instances reaching the dead zone, where dynamic temperature sampling results in a higher proportion.

undergoing temperature increases is higher compared to low sampling sizes, which aligns with our analysis presented in Section 3.

Proportion of Optimal Temperature Range We analyze the proportion of FSD instances that ultimately reach the dead zone. We consider reaching the dead zone as an indication that the sample operates within an optimal temperature range. As shown in Figure 9, dynamic temperature sampling results in a higher proportion of FSD instances entering the dead zone compared to fixed-temperature sampling, suggesting that our method enables better alignment for a larger number of samples.

6 Related Work

Self-Consistency Self-consistency (Wang et al., 2023), also known as majority voting, is a significant method for effectively enhancing the reasoning performance of large language models (LLMs) within the context of chain-of-thought (Wei et al., 2022) settings. Research on this method primarily focuses on two aspects: First, the effectiveness of self-consistency is further improved through weighted majority voting (Li et al., 2023a, 2024b) or input diversity (Sathe et al., 2024). Additionally, some have extended self-consistency to open-domain generation (Wang et al., 2024b; Jain et al., 2023), allowing its application beyond reasoning tasks. Second, some studies aim to reduce the cost of self-consistency without compromising performance, according to early stopping criteria about answer distributions (Li et al., 2024c; Aggarwal et al., 2023), difficulty (Wang et al., 2024a), quality (Wan et al., 2024) or consistency of reasoning paths (Zhu et al., 2024a). Chen et al. (2024) have employed a hybrid strategy combining sampling and greedy algorithms to reduce computational costs. Recently, theoretical analyses of voting strategies (Wu et al., 2024; Li et al., 2024c) were provided, offering a theoretical foundation for the study of self-consistency. Our method offers a deeper viewpoint, revisiting self-consistency from the perspective of distributional dynamic alignment.

Diversity Control for Language Models Decoding strategy is a critical factor in controlling the diversity of language models. From the perspective of the probability distribution of generated tokens, temperature sampling (Ackley et al., 1985) controls the sharpness of the distribution by adjusting the temperature. Existing research primarily focuses on diversity control within a single sampling process (Zhang et al., 2024; Zhu et al., 2024b; Dhuliawala et al., 2024; Li et al., 2024a). At the task level, Renze (2024) have examined the impact of temperature on the model's problem-solving capabilities. However, the influence of diversity control on self-consistency and the underlying mechanisms remain unexplored.

7 Conclusion

This work revisits self-consistency through the lens of dynamic distributional alignment, challenging the conventional view of passive convergence to a fixed answer distribution. We demonstrate that decoding temperature critically shapes both sampling behavior and the latent answer distribution itself, revealing a trade-off between diversity-driven exploration and finite-sample convergence. By introducing a confidence-aware mechanism that dynamically adjusts temperature based on real-time alignment with the distribution, we bridge this gap, enabling efficient synchronization between sampling dynamics and evolving answer distributions. Empirical results validate that this approach outperforms static strategies, achieving robust performance improvements without external resources. Our findings position self-consistency as an active alignment challenge, opening avenues for adaptive aggregation frameworks in reasoning tasks.

Limitations

While our approach advances the understanding and application of self-consistency, several limitations remain:

- Task Scope: Experiments focus on mathematical reasoning tasks, thus generalization to broader domains (e.g., open-ended generation or multi-step decision-making) requires further validation.
- Optimal Temperature: The specific value of the optimal temperature when the sample size approaches infinity, and how it varies with factors such as the model and dataset, remains unexplored.
- Decoding Strategy Interactions: The interplay between temperature modulation and other decoding techniques (e.g., top-k or top-p sampling) remains unexplored, potentially affecting broader applicability.

Ethics Statement

All of the datasets used in this study were publicly available, and no annotators were employed for our data collection. We confirm that the datasets we used did not contain any harmful content and was consistent with their intended use (research). We have cited the datasets and relevant works used in this study.

Acknowledgments

This work is supported by Beijing Natural Science Foundation (No.4222037, L181010).

References

David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. 1985. A learning algorithm for boltzmann machines. *Cogn. Sci.*, 9(1):147–169.

Pranjal Aggarwal, Aman Madaan, Yiming Yang, and Mausam. 2023. Let‘s sample step by step: Adaptive-consistency for efficient reasoning and coding with LLMs. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12375–12396, Singapore. Association for Computational Linguistics.

Wenqing Chen, Weicheng Wang, Zhixuan Chu, Kui Ren, Zibin Zheng, and Zhichao Lu. 2024. Self-para-consistency: Improving reasoning tasks at low cost for large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 14162–14167, Bangkok, Thailand. Association for Computational Linguistics.

Xinyun Chen, Renat Aksitov, Uri Alon, Jie Ren, Kefan Xiao, Pengcheng Yin, Sushant Prakash, Charles Sutton, Xuezhi Wang, and Denny Zhou. 2023. Universal self-consistency for large language model generation. *CoRR*, abs/2311.17311.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems. *CoRR*, abs/2110.14168.

Shehzaad Dhuliawala, Ilia Kulikov, Ping Yu, Asli Celikyilmaz, Jason Weston, Sainbayar Sukhbaatar, and Jack Lanchantin. 2024. Adaptive decoding via latent preference optimization. *CoRR*, abs/2411.09661.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurélien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Rozière, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Graeme Nail, Grégoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel M. Kloumann, Ishan Misra, Ivan Evtimov, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, and et al. 2024. The llama 3 herd of models. *CoRR*, abs/2407.21783.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the MATH dataset. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*.

Siddhartha Jain, Xiaofei Ma, Anoop Deoras, and Bing Xiang. 2023. Self-consistency for open-ended generations. *CoRR*, abs/2307.06857.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *CoRR*, abs/2310.06825.

Yifei Li, Zeqi Lin, Shizhuo Zhang, Qiang Fu, Bei Chen, Jian-Guang Lou, and Weizhu Chen. 2023a. Making language models better reasoners with step-aware verifier. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5315–5333, Toronto, Canada. Association for Computational Linguistics.

Yiwei Li, Fei Mi, Yitong Li, Yasheng Wang, Bin Sun, Shaoxiong Feng, and Kan Li. 2024a. Dynamic stochastic decoding strategy for open-domain dialogue generation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 11585–11596, Bangkok, Thailand. Association for Computational Linguistics.

Yiwei Li, Peiwen Yuan, Shaoxiong Feng, Boyuan Pan, Bin Sun, Xinglin Wang, Heda Wang, and Kan Li. 2024b. Turning dust into gold: Distilling complex reasoning capabilities from llms by leveraging negative data. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 18591–18599. AAAI Press.

Yiwei Li, Peiwen Yuan, Shaoxiong Feng, Boyuan Pan, Xinglin Wang, Bin Sun, Heda Wang, and Kan Li. 2024c. Escape sky-high cost: Early-stopping self-consistency for multi-step reasoning. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Yuanzhi Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. 2023b. Textbooks are all you need II: phi-1.5 technical report. *CoRR*, abs/2309.05463.

Qing Lyu, Kumar Shridhar, Chaitanya Malaviya, Li Zhang, Yanai Elazar, Niket Tandon, Marianna Apidianaki, Mrinmaya Sachan, and Chris Callison-Burch. 2024. Calibrating large language models with sample consistency. *CoRR*, abs/2402.13904.

Matthew Renze. 2024. The effect of sampling temperature on problem solving in large language models. In *Findings of the Association for Computational*

Linguistics: EMNLP 2024, pages 7346–7356, Miami, Florida, USA. Association for Computational Linguistics.

Morgane Rivière, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, Anton Tsitsulin, Nino Vieillard, Piotr Stanczyk, Sertan Girgin, Nikola Momchev, Matt Hoffman, Shantanu Thakoor, Jean-Bastien Grill, Behnam Neyshabur, Olivier Bachem, Alanna Walton, Aliaksei Severyn, Alicia Parrish, Aliya Ahmad, Allen Hutchison, Alvin Abdagic, Amanda Carl, Amy Shen, Andy Brock, Andy Coenen, Anthony Laforge, Antonia Paterson, Ben Bastian, Bilal Piot, Bo Wu, Brandon Royal, Charlie Chen, Chintu Kumar, Chris Perry, Chris Welty, Christopher A. Choquette-Choo, Danila Sinopalnikov, David Weinberger, Dimple Vijaykumar, Dominika Rogozinska, Dustin Herbison, Elisa Bandy, Emma Wang, Eric Noland, Erica Moreira, Evan Senter, Evgenii Eltyshev, Francesco Visin, Gabriel Rasskin, Gary Wei, Glenn Cameron, Gus Martins, Hadi Hashemi, Hanna Klimczak-Plucinska, Harleen Batra, Harsh Dhand, Ivan Nardini, Jacinda Mein, Jack Zhou, James Svensson, Jeff Stanway, Jetha Chan, Jin Peng Zhou, Joana Carrasqueira, Joana Iljazi, Jocelyn Becker, Joe Fernandez, Joost van Amersfoort, Josh Gordon, Josh Lipschultz, Josh Newlan, Ju-yeong Ji, Kareem Mohamed, Kartikeya Badola, Kat Black, Katie Millican, Keelin McDonell, Kelvin Nguyen, Kiranbir Sodhia, Kish Greene, Lars Lowe Sjösund, Lauren Usui, Laurent Sifre, Lena Heuermann, Leticia Lago, and Lilly McNealus. 2024. Gemma 2: Improving open language models at a practical size. *CoRR*, abs/2408.00118.

Ashutosh Sathe, Divyanshu Aggarwal, and Sunayana Sitaram. 2024. Improving self consistency in llms through probabilistic tokenization. *CoRR*, abs/2407.03678.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *CoRR*, abs/2402.03300.

Guangya Wan, Yuqi Wu, Jie Chen, and Sheng Li. 2024. Dynamic self-consistency: Leveraging reasoning paths for efficient LLM sampling. *CoRR*, abs/2408.17017.

Xinglin Wang, Shaoxiong Feng, Yiwei Li, Peiwen Yuan, Yueqi Zhang, Boyuan Pan, Heda Wang, Yao Hu, and Kan Li. 2024a. Make every penny count: Difficulty-adaptive self-consistency for cost-efficient reasoning. *CoRR*, abs/2408.13457.

Xinglin Wang, Yiwei Li, Shaoxiong Feng, Peiwen Yuan, Boyuan Pan, Heda Wang, Yao Hu, and Kan Li. 2024b. Integrate the essence and eliminate the dross: Fine-grained self-consistency for free-form language generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11782–11794, Bangkok, Thailand. Association for Computational Linguistics.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *NeurIPS*.

Yangzhen Wu, Zhiqing Sun, Shanda Li, Sean Welleck, and Yiming Yang. 2024. Inference scaling laws: An empirical analysis of compute-optimal inference for llm problem-solving. In *The 4th Workshop on Mathematical Reasoning and AI at NeurIPS'24*.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. 2024. Qwen2.5 technical report. *CoRR*, abs/2412.15115.

Shimao Zhang, Yu Bao, and Shujian Huang. 2024. EDT: improving large language models' generation by entropy-based dynamic temperature sampling. *CoRR*, abs/2403.14541.

Jiace Zhu, Yingtao Shen, Jie Zhao, and An Zou. 2024a. Path-consistency: Prefix enhancement for efficient inference in LLM. *CoRR*, abs/2409.01281.

Yuqi Zhu, Jia Li, Ge Li, Yunfei Zhao, Jia Li, Zhi Jin, and Hong Mei. 2024b. Hot or cold? adaptive temperature sampling for code generation with large language models. In *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 437–445. AAAI Press.

A Proof of Theorem 2.2

Proof A.1. *Firstly, we need to introduce true labels into Definition 2.1. As we are not concerned with the specific numerical values of the predicted and true answers, we map the set of predicted answers onto a sequence of natural numbers (in simple terms, we only need to know which of the i-th answers is the correct one). Consequently, we can establish the following partial order relation:*

$$\begin{aligned}
&\left| \underset{i}{argmax}\ f^M(i) - \underset{i}{argmax}\ f^N(i) \right| \\
&= \left| [\underset{i}{argmax}\ f^M(i) - gt_j] \right. \\
&\qquad \left. -[\underset{i}{argmax}\ f^N(i) - gt_j] \right| \\
&\geq \left| \mathbb{I}[\underset{i}{argmax}\ f^M(i) = gt_j] \right. \\
&\qquad \left. -\mathbb{I}[\underset{i}{argmax}\ f^N(i) = gt_j] \right|
\end{aligned} \tag{15}$$

Based on Definition 2.1, we have:

$$\begin{aligned}
&\left| \mathbb{I}[\underset{i}{argmax}\ f^M(i) = gt_j] \right. \\
&\qquad \left. -\mathbb{I}[\underset{i}{argmax}\ f^N(i) = gt_j] \right| < \epsilon
\end{aligned} \tag{16}$$

Next, we introduce the dataset D into Equation 16:

$$\begin{aligned}
&\frac{1}{|D|}\sum_{j\in D} \left| \mathbb{I}[\underset{i}{argmax}\ f^M(i) = gt_j] \right. \\
&\qquad \left. -\mathbb{I}[\underset{i}{argmax}\ f^N(i) = gt_j] \right| < \epsilon
\end{aligned} \tag{17}$$

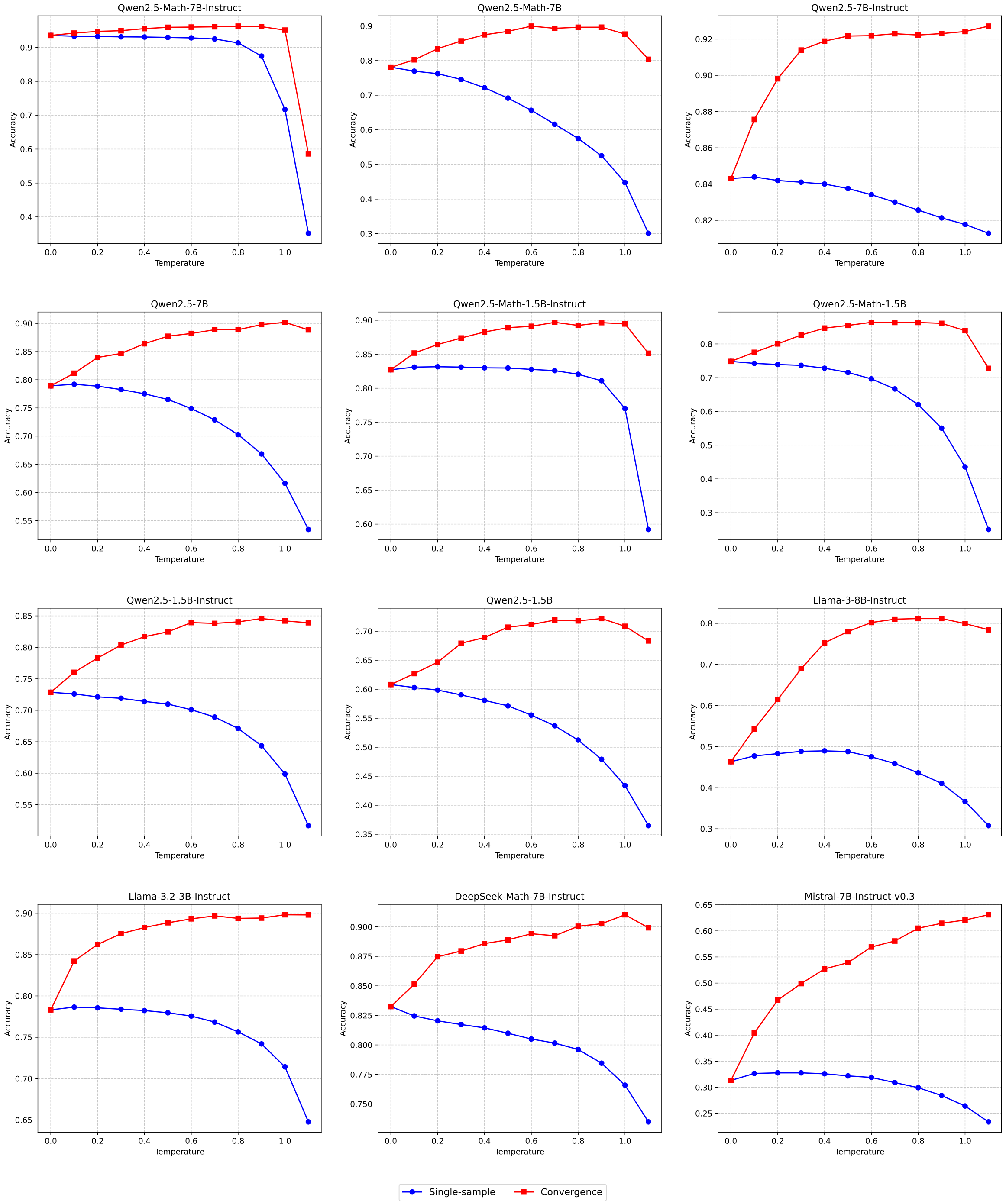
According to $|a+b| \leq |a| + |b|$, we have:

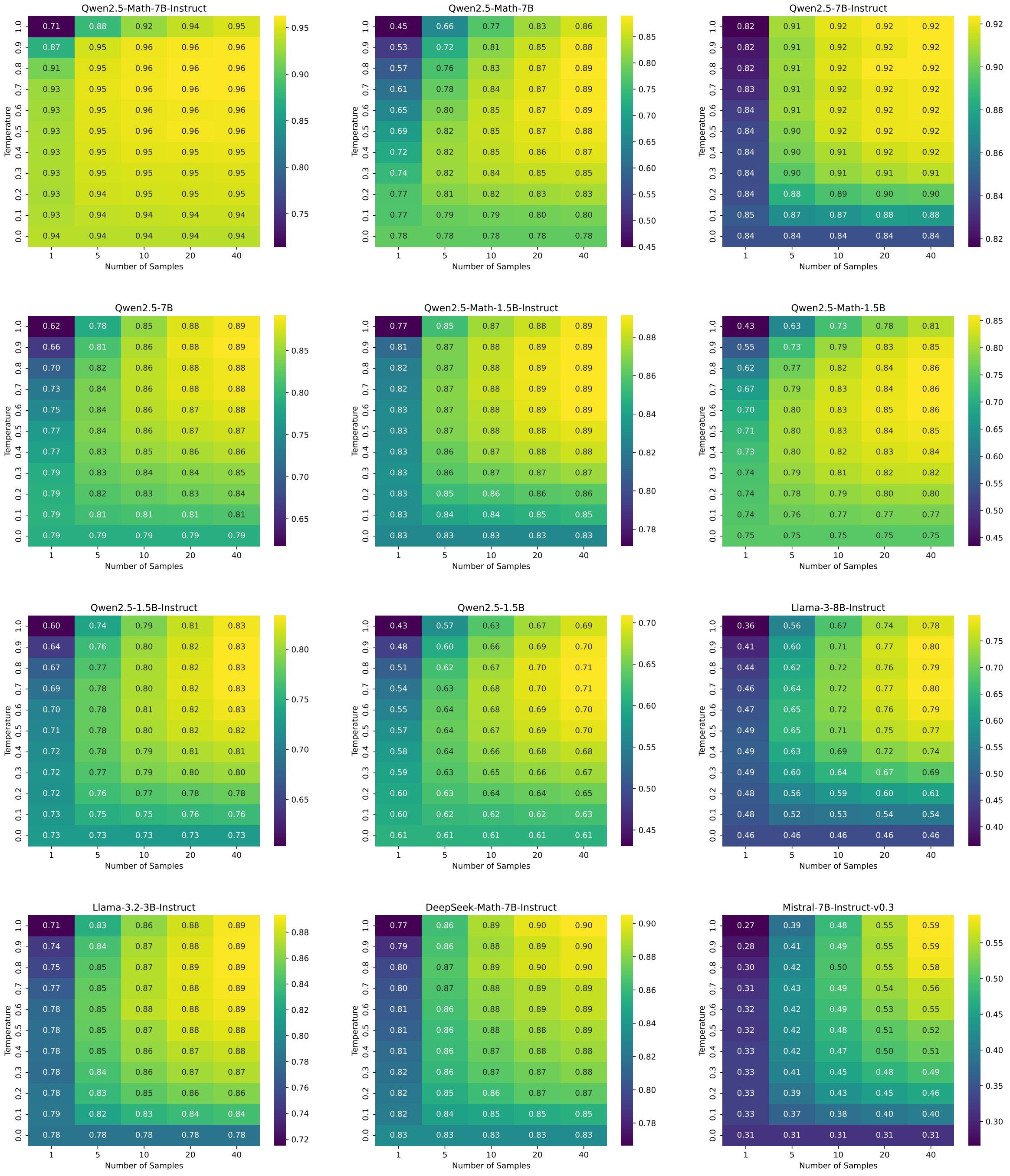
$$\begin{aligned}
&\left| \frac{1}{|D|}\sum_{j\in D} \mathbb{I}[\underset{i}{argmax}\ f^M(i) = gt_j] \right. \\
&\left. -\frac{1}{|D|}\sum_{j\in D} \mathbb{I}[\underset{i}{argmax}\ f^N(i) = gt_j] \right| < \epsilon
\end{aligned} \tag{18}$$

Finally, we can derive Theorem 2.2:

$$\left|\ Acc_D^M - Acc_D^N\ \right| < \epsilon \tag{19}$$

B Additional Results of Section 3





Qwen2.5-Math-7B-Instruct

Temperature	1	5	10	20	40
1.0	0.71	0.88	0.92	0.94	0.95
0.9	0.87	0.95	0.96	0.96	0.96
0.8	0.91	0.95	0.96	0.96	0.96
0.7	0.93	0.95	0.96	0.96	0.96
0.6	0.93	0.95	0.96	0.96	0.96
0.5	0.93	0.95	0.96	0.96	0.96
0.4	0.93	0.95	0.95	0.95	0.95
0.3	0.93	0.95	0.95	0.95	0.95
0.2	0.93	0.94	0.95	0.95	0.95
0.1	0.93	0.94	0.94	0.94	0.94
0.0	0.94	0.94	0.94	0.94	0.94

Number of Samples

Qwen2.5-Math-7B

Temperature	1	5	10	20	40
1.0	0.45	0.66	0.77	0.83	0.86
0.9	0.53	0.72	0.81	0.85	0.88
0.8	0.57	0.76	0.83	0.87	0.89
0.7	0.61	0.78	0.84	0.87	0.89
0.6	0.65	0.80	0.85	0.87	0.89
0.5	0.69	0.82	0.85	0.87	0.88
0.4	0.72	0.82	0.85	0.86	0.87
0.3	0.74	0.82	0.84	0.85	0.85
0.2	0.77	0.81	0.82	0.83	0.83
0.1	0.77	0.79	0.79	0.80	0.80
0.0	0.78	0.78	0.78	0.78	0.78

Number of Samples

Qwen2.5-7B-Instruct

Temperature	1	5	10	20	40
1.0	0.82	0.91	0.92	0.92	0.92
0.9	0.82	0.91	0.92	0.92	0.92
0.8	0.82	0.91	0.92	0.92	0.92
0.7	0.83	0.91	0.92	0.92	0.92
0.6	0.84	0.91	0.92	0.92	0.92
0.5	0.84	0.90	0.92	0.92	0.92
0.4	0.84	0.90	0.91	0.92	0.92
0.3	0.84	0.90	0.91	0.91	0.91
0.2	0.84	0.88	0.89	0.90	0.90
0.1	0.85	0.87	0.87	0.88	0.88
0.0	0.84	0.84	0.84	0.84	0.84

Number of Samples

Qwen2.5-7B

Temperature	1	5	10	20	40
1.0	0.62	0.78	0.85	0.88	0.89
0.9	0.66	0.81	0.86	0.88	0.89
0.8	0.70	0.82	0.86	0.88	0.88
0.7	0.73	0.84	0.86	0.88	0.88
0.6	0.75	0.84	0.86	0.87	0.88
0.5	0.77	0.84	0.86	0.87	0.87
0.4	0.77	0.83	0.85	0.86	0.86
0.3	0.79	0.83	0.84	0.84	0.85
0.2	0.79	0.82	0.83	0.83	0.84
0.1	0.79	0.81	0.81	0.81	0.81
0.0	0.79	0.79	0.79	0.79	0.79

Number of Samples

Qwen2.5-Math-1.5B-Instruct

Temperature	1	5	10	20	40
1.0	0.77	0.85	0.87	0.88	0.89
0.9	0.81	0.87	0.88	0.89	0.89
0.8	0.82	0.87	0.88	0.89	0.89
0.7	0.82	0.87	0.88	0.89	0.89
0.6	0.83	0.87	0.88	0.89	0.89
0.5	0.83	0.87	0.88	0.88	0.89
0.4	0.83	0.86	0.87	0.88	0.88
0.3	0.83	0.86	0.87	0.87	0.87
0.2	0.83	0.85	0.86	0.86	0.86
0.1	0.83	0.84	0.84	0.85	0.85
0.0	0.83	0.83	0.83	0.83	0.83

Number of Samples

Qwen2.5-Math-1.5B

Temperature	1	5	10	20	40
1.0	0.43	0.63	0.73	0.78	0.81
0.9	0.55	0.73	0.79	0.83	0.85
0.8	0.62	0.77	0.82	0.84	0.86
0.7	0.67	0.79	0.83	0.84	0.86
0.6	0.70	0.80	0.83	0.85	0.86
0.5	0.71	0.80	0.83	0.84	0.85
0.4	0.73	0.80	0.82	0.83	0.84
0.3	0.74	0.79	0.81	0.82	0.82
0.2	0.74	0.78	0.79	0.80	0.80
0.1	0.74	0.76	0.77	0.77	0.77
0.0	0.75	0.75	0.75	0.75	0.75

Number of Samples

Qwen2.5-1.5B-Instruct

Temperature	1	5	10	20	40
1.0	0.60	0.74	0.79	0.81	0.83
0.9	0.64	0.76	0.80	0.82	0.83
0.8	0.67	0.77	0.80	0.82	0.83
0.7	0.69	0.78	0.80	0.82	0.83
0.6	0.70	0.78	0.81	0.82	0.83
0.5	0.71	0.78	0.80	0.82	0.82
0.4	0.72	0.78	0.79	0.81	0.81
0.3	0.72	0.77	0.79	0.80	0.80
0.2	0.72	0.76	0.77	0.78	0.78
0.1	0.73	0.75	0.75	0.76	0.76
0.0	0.73	0.73	0.73	0.73	0.73

Number of Samples

Qwen2.5-1.5B

Temperature	1	5	10	20	40
1.0	0.43	0.57	0.63	0.67	0.69
0.9	0.48	0.60	0.66	0.69	0.70
0.8	0.51	0.62	0.67	0.70	0.71
0.7	0.54	0.63	0.68	0.70	0.71
0.6	0.55	0.64	0.68	0.69	0.70
0.5	0.57	0.64	0.67	0.69	0.70
0.4	0.58	0.64	0.66	0.68	0.68
0.3	0.59	0.63	0.65	0.66	0.67
0.2	0.60	0.63	0.64	0.64	0.65
0.1	0.60	0.62	0.62	0.62	0.63
0.0	0.61	0.61	0.61	0.61	0.61

Number of Samples

Llama-3-8B-Instruct

Temperature	1	5	10	20	40
1.0	0.36	0.56	0.67	0.74	0.78
0.9	0.41	0.60	0.71	0.77	0.80
0.8	0.44	0.62	0.72	0.76	0.79
0.7	0.46	0.64	0.72	0.77	0.80
0.6	0.47	0.65	0.72	0.76	0.79
0.5	0.49	0.65	0.71	0.75	0.77
0.4	0.49	0.63	0.69	0.72	0.74
0.3	0.49	0.60	0.64	0.67	0.69
0.2	0.48	0.56	0.59	0.60	0.61
0.1	0.48	0.52	0.53	0.54	0.54
0.0	0.46	0.46	0.46	0.46	0.46

Number of Samples

Llama-3.2-3B-Instruct

Temperature	1	5	10	20	40
1.0	0.71	0.83	0.86	0.88	0.89
0.9	0.74	0.84	0.87	0.88	0.89
0.8	0.75	0.85	0.87	0.89	0.89
0.7	0.77	0.85	0.87	0.88	0.89
0.6	0.78	0.85	0.88	0.88	0.89
0.5	0.78	0.85	0.87	0.88	0.88
0.4	0.78	0.85	0.86	0.87	0.88
0.3	0.78	0.84	0.86	0.87	0.87
0.2	0.78	0.83	0.85	0.86	0.86
0.1	0.79	0.82	0.83	0.84	0.84
0.0	0.78	0.78	0.78	0.78	0.78

Number of Samples

DeepSeek-Math-7B-Instruct

Temperature	1	5	10	20	40
1.0	0.77	0.86	0.89	0.90	0.90
0.9	0.79	0.86	0.88	0.89	0.90
0.8	0.80	0.87	0.89	0.90	0.90
0.7	0.80	0.87	0.88	0.89	0.89
0.6	0.81	0.86	0.88	0.89	0.89
0.5	0.81	0.86	0.88	0.88	0.89
0.4	0.81	0.86	0.87	0.88	0.88
0.3	0.82	0.86	0.87	0.87	0.88
0.2	0.82	0.85	0.86	0.87	0.87
0.1	0.82	0.84	0.85	0.85	0.85
0.0	0.83	0.83	0.83	0.83	0.83

Number of Samples

Mistral-7B-Instruct-v0.3

Temperature	1	5	10	20	40
1.0	0.27	0.39	0.48	0.55	0.59
0.9	0.28	0.41	0.49	0.55	0.59
0.8	0.30	0.42	0.50	0.55	0.58
0.7	0.31	0.43	0.49	0.54	0.56
0.6	0.32	0.42	0.49	0.53	0.55
0.5	0.32	0.42	0.48	0.51	0.52
0.4	0.33	0.42	0.47	0.50	0.51
0.3	0.33	0.41	0.45	0.48	0.49
0.2	0.33	0.39	0.43	0.45	0.46
0.1	0.33	0.37	0.38	0.40	0.40
0.0	0.31	0.31	0.31	0.31	0.31

Number of Samples

C Justification of the Top-2 Assumption

To validate the correctness of the assumption in Section 4.3 that the final answer can only appear among the top-2 answers, we conducted an analysis to measure how often the final self-consistency answer overlaps with the top-2 answers observed in early samples.

Model	$N = 10$	$N = 20$
Qwen2.5-Math-7B-Instruct	99.44%	99.73%
Qwen2.5-Math-7B	96.79%	98.39%
Qwen2.5-7B	97.74%	98.82%
Qwen2.5-1.5B-Instruct	95.38%	97.53%
Qwen2.5-7B-Instruct	99.21%	99.58%
LLaMA-3-8B-Instruct	87.85%	93.65%
LLaMA-3.2-3B-Instruct	96.86%	98.31%

Table 2: Overlap Rate Between Top-2 and Final Answer.

The results in Table 2 suggest that the final majority-vote answer is overwhelmingly likely to appear among the top-2 candidates, even with a small number of samples. This supports the practical validity of the assumption.

D Additional Results of Section 5.2

The results in Table 3 demonstrate that our method performs well across reasoning tasks in different domains.

E Additional Analysis of Section 5.3

E.1 Influence of Model Architecture and Calibration Properties

The underlying model architecture can influence the effectiveness of self-consistency, particularly due to differences in reasoning ability, calibration behavior, and sensitivity to temperature. To explore this, we conducted a comparative analysis across different backbone models, using several indicators: (1) Confidence is measured via answer entropy and FSD. (2) Stability is measured via the variance (Var) of accuracy under different fixed temperatures. (3) Effectiveness is reflected by both the absolute accuracy under fixed-temperature self-consistency (Acc fix@N) and the performance gains brought by our adaptive method (Gain@N).

According to Table 4, our key observation is that higher confidence models (lower entropy, higher FSD) tend to: (1) achieve higher base accuracy under fixed temperature, and (2) exhibit lower sensitivity to temperature (i.e. lower variance), resulting in smaller performance gains from adaptive strategies. These trends align well with our understanding of model behavior: stronger models tend to produce more confident predictions, making them inherently less reliant on temperature-based sampling adjustments. Conversely, less confident models benefit more from dynamic temperature calibration, as their sampling distributions are more sensitive to the choice of temperature.

E.2 Relationship Between Sample Difficulty and Temperature Variation

To better understand the reasons behind the differing behavior of the observed temperature variation between samples, we hypothesize that sample difficulty is a key prior factor. Intuitively, harder questions tend to result in lower model confidence and may require lower initial temperatures to guide convergence, whereas easier questions are more stable and better explored with higher temperatures.

To examine this hypothesis, we conducted further analysis using samples with known or estimated difficulty levels. For the MATH dataset, we use its ground-truth difficulty labels (1–5). For GSM8K, we used an LLM-based difficulty estimation strategy, where we applied repeated batch-wise comparisons to assign continuous difficulty scores (1–8).

For each initial temperature T_0, we group the samples based on whether the final temperature increases, decreases, or remains the same, and then calculate the average difficulty score within each group.

The results in Table 5 show a clear and consistent pattern: (1) More difficult questions tend to lead to temperature decreases, while easier questions often allow for temperature increases. (2) Higher initial temperatures generally result in more samples decreasing their temperature during the adaptive process. This aligns well with our intuition and further supports the idea that the final temperature T_3 is implicitly influenced by sample difficulty.

Model	Strategy	Last Letter Concatenation						StrategyQA					
		N=10		N=20		N=40		N=10		N=20		N=40	
		Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Qwen2.5-1.5B-Instruct	Fix	12.7	16.6	15.6	19.5	18.4	22.2	55.1	58.9	57.4	60.5	58.7	61.2
	Dynamic	**13.7**	**16.8**	**18.5**	**20.9**	**21.5**	**23.3**	**55.8**	**59.1**	**58.2**	**60.7**	**59.5**	**61.4**
LLaMA-3.2-3B-Instruct	Fix	72.8	76.1	75.7	78.9	76.7	80.2	67.1	**70.5**	67.9	**71.3**	68.3	**71.5**
	Dynamic	**73.4**	**76.5**	**76.6**	**79.4**	**78.2**	**81.0**	**68.2**	**70.5**	**68.8**	**71.3**	**68.9**	**71.5**
Mistral-7B-Instruct-v0.3	Fix	4.8	5.7	6.0	7.9	**6.8**	9.5	52.3	55.0	55.2	59.9	56.6	62.3
	Dynamic	**5.1**	**6.1**	**6.2**	**8.9**	**6.8**	**10.3**	**52.6**	**55.9**	**55.8**	**60.7**	**57.5**	**62.9**

Table 3: Evaluation results on Last Letter Concatenation and StrategyQA tasks.

Model	Entropy	FSD	Var	Acc@10	Gain@10	Acc@20	Gain@20	Acc@40	Gain@40
Qwen2.5-1.5B-Instruct	1.97	0.483	4.089	79.0	+0.2	80.3	+0.5	81.1	+0.5
Qwen2.5-7B-Instruct	1.17	0.656	2.120	90.8	+0.0	91.2	+0.2	91.4	+0.3
LLaMA-3.2-3B-Instruct	1.96	0.526	2.146	86.2	+0.0	87.2	+0.3	87.7	+0.4
LLaMA-3-8B-Instruct	2.61	0.349	52.387	66.6	+0.5	70.2	+1.4	76.1	+1.9

Table 4: Comparison of model uncertainty, stability, and effectiveness under fixed-temperature and adaptive method.

T_0	GSM8K			MATH		
	$\uparrow T_3$ Avg Level	$\rightarrow T_3$ Avg Level	$\downarrow T_3$ Avg Level	$\uparrow T_3$ Avg Level	$\rightarrow T_3$ Avg Level	$\downarrow T_3$ Avg Level
0.2	4.41	5.27	6.06	3.29	4.08	4.21
0.3	4.43	5.34	5.90	3.19	4.05	4.16
0.4	4.24	5.44	5.63	3.03	3.95	4.10
0.5	4.22	5.27	5.67	3.00	3.86	4.10
0.6	4.20	5.25	5.70	2.99	3.81	4.10
0.7	4.20	5.24	5.65	2.98	3.78	4.09
0.8	4.16	5.20	5.61	2.94	3.71	4.10
0.9	4.12	5.09	5.66	2.91	3.73	4.04

Table 5: Average sample difficulty levels by temperature adaptation results across different initial temperatures for GSM8K and MATH.