

C.8. Temperature and Entropy Analysis

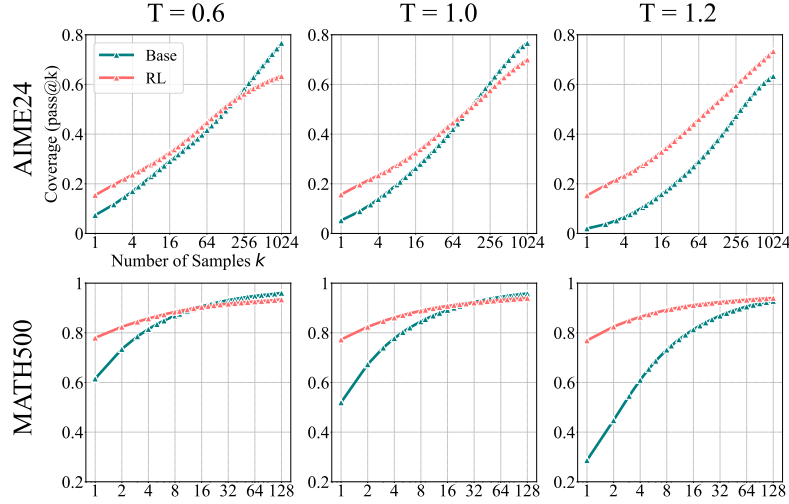


Figure 17: We found that the base model’s performance drops when the temperature exceeds 1.0, as it tends to generate more random and less coherent tokens. In contrast, the RL model’s performance remains relatively stable across different temperature settings. Therefore, we use $T = 0.6$ in the main experiments, as it allows both models to demonstrate their best reasoning performance.

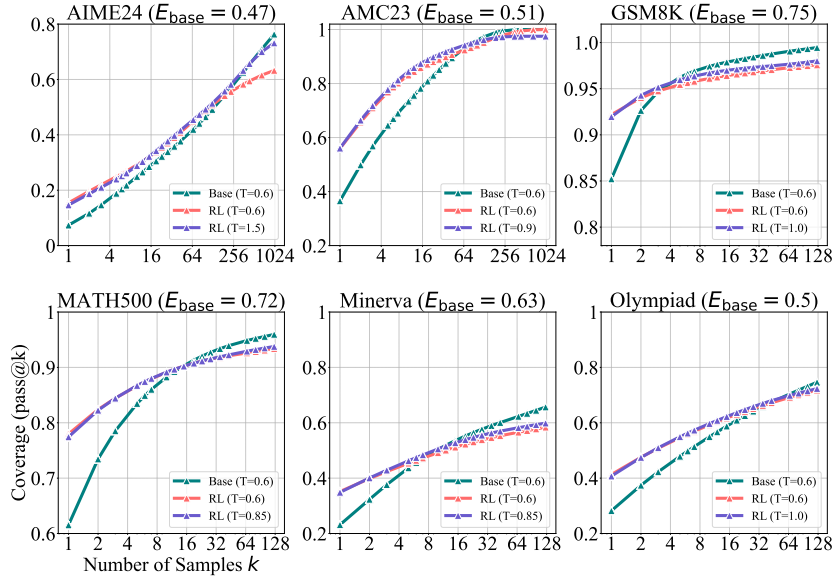


Figure 18: **Comparison of Base and RLVR Models with Matched Output Entropy.** We evaluate the base model (Qwen2.5-7B) on each dataset using temperature $T = 0.6$ and report its output entropy E_{base} in the title of each figure. To enable a fair comparison, we increase the temperature of the RLVR model (SimpleRLZoo) until its output entropy approximately matches E_{base} . For example, on AMC23, we set $T = 0.9$ to achieve $E_{\text{RL}} = 0.47$. We also include RLVR results at $T = 0.6$ as an additional baseline, which has lower entropy—e.g., 0.22 on AMC23 and 0.33 on MATH500.

C.9. Training Dynamics

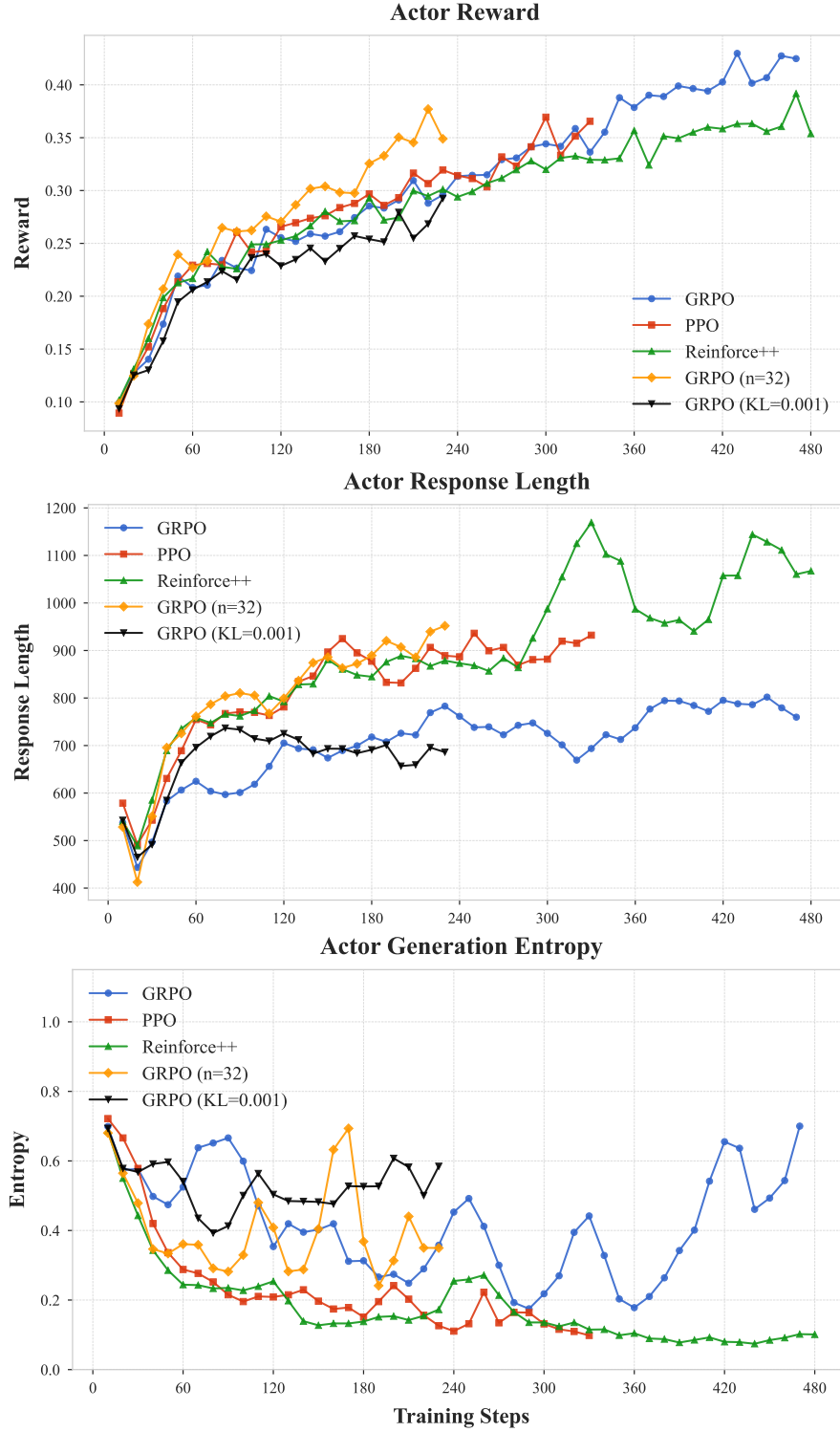


Figure 19: The curves of training reward, response length, and generation entropy during training, corresponding to experiments in Section 4.

C.10. CoT Case Analysis

Question

(AIME24 Question No.16)

Among the 900 residents of Aimeville, there are 195 who own a diamond ring, 367 who own a set of golf clubs, and 562 who own a garden spade. In addition, each of the 900 residents owns a bag of candy hearts. There are 437 residents who own exactly two of these things, and 234 residents who own exactly three of these things. Find the number of residents of Aimeville who own all four of these things.

Qwen-2.5-7B Answer

Response Length: 1058

To solve this problem, we can use the principle of inclusion-exclusion. We will denote the number of residents who own diamond rings, golf clubs, garden spades, and candy hearts as D, G, S, and C, respectively. We are given the following information:

- D = 195
- G = 367
- S = 562
- C = 900 (since every resident owns a bag of candy hearts)

We are also given that 437 residents own exactly two of these things and 234 residents own exactly three of these things.

Let's denote the number of residents who own all four of these things as X. We need to find X.

First, we need to find the number of residents who own at least one of these things. This can be calculated using the principle of inclusion-exclusion:

$$\begin{aligned} \text{Total} &= D + G + S + C - (\text{number of residents who own exactly two things}) \\ &\quad - 2 \times (\text{number of residents who own exactly three things}) - 3 \times (\text{number of residents who own all four things}) \end{aligned}$$

Since there are 900 residents in total, we can write the equation as:

$$900 = 195 + 367 + 562 + 900 - 437 - 2 \times 234 - 3 \times X$$

Now we need to solve for X:

$$\begin{aligned} 900 &= 195 + 367 + 562 + 900 - 437 - 468 - 3 \times X \\ 900 &= 1717 - 905 - 3 \times X \\ 900 &= 812 - 3 \times X \\ 3 \times X &= 812 - 900 \\ 3 \times X &= -88 \\ X &= -\frac{88}{3} \end{aligned}$$

Since the number of residents cannot be negative, there is an error in our calculations. Let's re-evaluate the problem.

We know that there are 437 residents who own exactly two things and 234 residents who own exactly three things. We can subtract these from the total number of residents who own at least one thing to find the number of residents who own exactly one thing:

$$\begin{aligned} \text{Number of residents who own exactly one thing} &= (D + G + S + C) - (\text{number of residents who own exactly two things}) \\ &\quad - 2 \times (\text{number of residents who own exactly three things}) - 3 \times (\text{number of residents who own all four things}) \end{aligned}$$

Let's denote the number of residents who own exactly one thing as Y:

$$\begin{aligned} Y &= (195 + 367 + 562 + 900) - 437 - 2 \times 234 - 3 \times X \\ Y &= 1924 - 905 - 3 \times X \\ Y &= 1019 - 3 \times X \end{aligned}$$

Since every resident owns at least one thing, we can write the equation:

$$900 = Y + 437 + 2 \times 234 + 3 \times X$$

Now we need to solve for X:

$$\begin{aligned} 900 &= (1019 - 3 \times X) + 437 + 468 + 3 \times X \\ 900 &= 1019 + 437 + 468 \\ 900 &= 1924 \end{aligned}$$

This equation is not true, which means there is still an error in our calculations. Let's try a different approach.

We can use the fact that the total number of residents who own at least one thing is 900. We can write the equation:

$$\begin{aligned} 900 &= (\text{number of residents who own exactly one thing}) + (\text{number of residents who own exactly two things}) \\ &\quad + (\text{number of residents who own exactly three things}) + (\text{number of residents who own all four things}) \end{aligned}$$

Let's denote the number of residents who own exactly one thing as Y:

$$\begin{aligned} 900 &= Y + 437 + 2 \times 234 + X \\ 900 &= Y + 805 + X \end{aligned}$$

Now we need to find Y. We can use the fact that the total number of residents who own at least one thing is equal to the sum of the residents who own each item minus the residents who own at least two items:

$$\begin{aligned} 900 &= (195 + 367 + 562 + 900) - (\text{number of residents who own exactly two things}) \\ &\quad - 2 \times (\text{number of residents who own exactly three things}) - 3 \times (\text{number of residents who own all four things}) \\ 900 &= 1924 - 437 - 2 \times 234 - 3 \times X \\ 900 &= 1924 - 805 - 3 \times X \\ 900 &= 1119 - 3 \times X \end{aligned}$$

Now we can solve for X:

$$\begin{aligned} 3 \times X &= 1119 - 900 \\ 3 \times X &= 219 \\ X &= 73 \end{aligned}$$

So, the number of residents of Aimeville who own all four of these things is 73.

Figure 20: Qwen2.5-Base-7B Correct Response - Case 1.