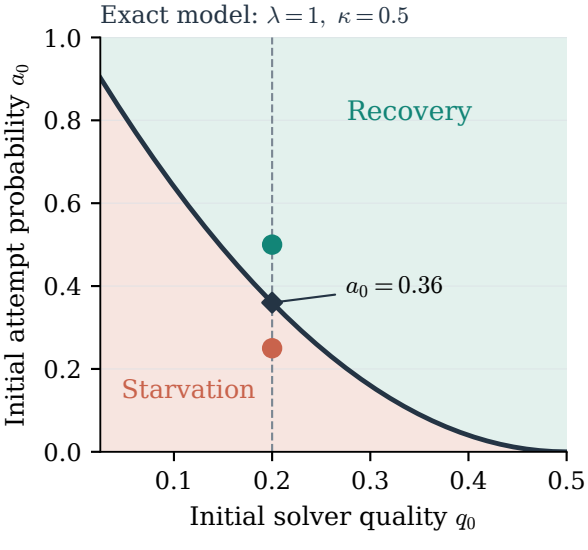
(a) Attempt starvation



(b) Two observed score decisions

Qwen3-8B · Binary RL · seed 41 · 4K

Case 5 · Omni-MATH — **CORRECT**

Output	$-(2+\sqrt{3})$	Reference	$-2-\sqrt{3}$
Box NLL	0.643	**Reject correct**	
Tail NLL	1.286	**Accept correct**	

Case 10 · DAPO — **WRONG**

Output	3	Reference	6
Box NLL	8.391×10^{-7}	**Accept wrong**	
Tail NLL	2.225	**Reject wrong**	

DAPO cutoffs: Box 2.208×10^{-6}; Tail 1.837

Accept if score $<$ cutoff; $\alpha = 0.065$