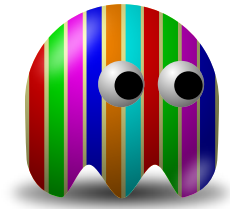


Shifting Experts on Easy Data



Manfred K. Warmuth

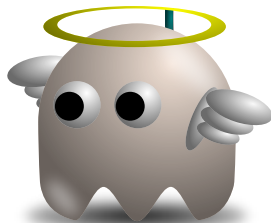
Wouter M. Koolen



Eternal Dilemma



Worst Case



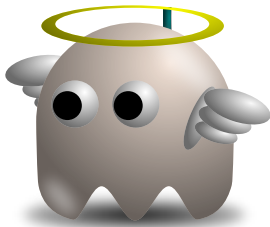
IID

Huge Difference (Expert Setting Example)



Hedge

$$O(\sqrt{T})$$



FTL

$$O(1)$$

Holy Grail



Adaptive

Status Quo



Experts	Hedge	FTL/ERM	FlipFlop
Bandits	EXP3	UCB	SAO
Shifting	Fixed Share	?	?
Freund's Problem	Mixing Past Posteriors	?	?

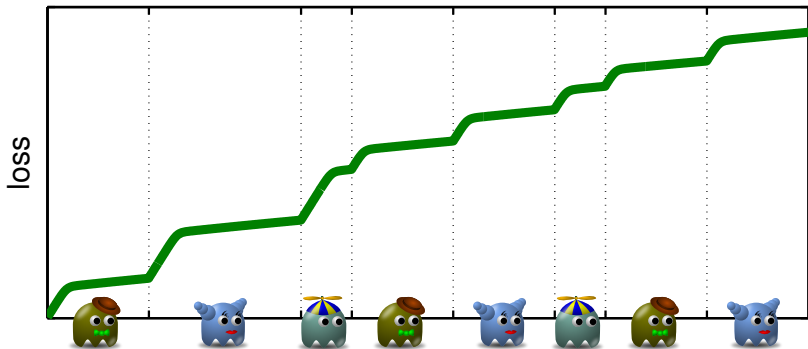
Shifting

Best model (expert) changes over time.



- ▶ Optimal algorithm for segment-wise IID data?
Should pay $O(\ln \# \text{experts})$ per switch
- ▶ How to combine it with worst-case robustness?

Freund's Problem: Long-term Memory



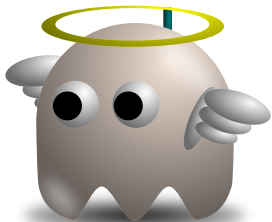
- ▶ Optimal algorithm for segment-wise IID data?
Should pay $O(\ln \# \text{good experts})$ per switch
- ▶ How to combine it with worst-case robustness?

Candidate Algorithms

For IID shifting:

- ▶ FL on the best partition
- ▶ FL on a shifting window
- ▶ FL on capped loss differences
- ▶ FL on exponentially decaying losses
- ▶ ...

For IID long-term memory?



The Big Question

Single algorithm for shifting

- ▶ worst-case robust
- ▶ adaptive to IID data

