

# **Multi-Objective Online Learning**

Chen-Yu Wei

# References

Chapter 3 of Aaron's lecture notes:

<https://www.cis.upenn.edu/~aaroath/GamesInLearning.pdf>

# **Group-wise Regret**

# Online Learning with Multiple Subsequences

For  $t = 1, 2, \dots, T$ :

Learner observes a context  $C_t \in \{0, 1\}^K$

Learner generates a distribution  $p_t \in \Delta_N$

Adversary decides the loss vector  $\ell_t = (\ell_t(1), \dots, \ell_t(N)) \in \mathbb{R}^N$

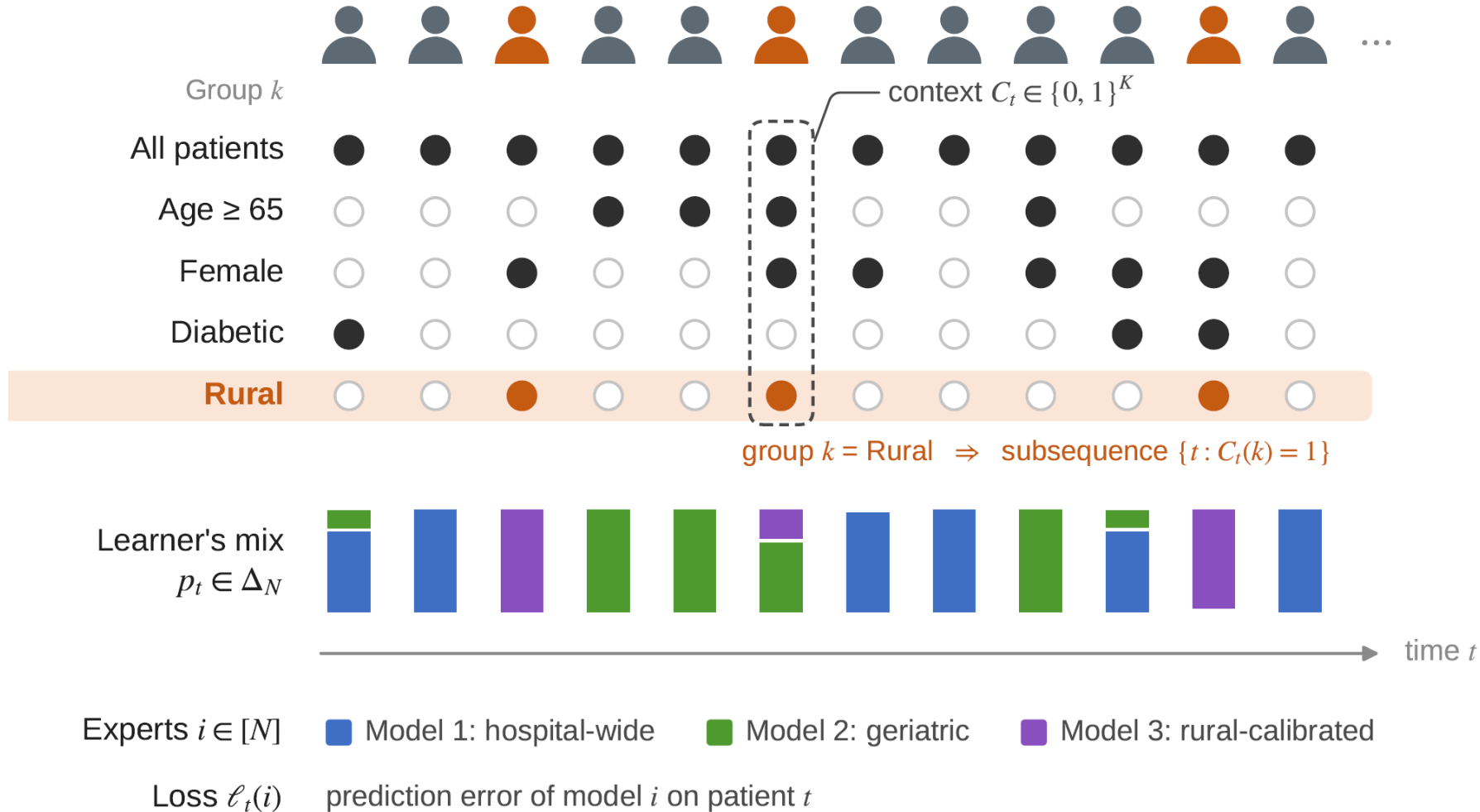
Adversary reveals  $\ell_t$  to the learner

Learner suffers loss  $\langle p_t, \ell_t \rangle$

$$\begin{aligned} \text{Group-wise Regret} &= \max_{k \in [K]} \left( \sum_{t: C_t(k)=1} \langle p_t, \ell_t \rangle - \min_{i \in [N]} \sum_{t: C_t(k)=1} \ell_t(i) \right) \\ &= \max_{k \in [K]} \max_{i \in [N]} \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(i)) \end{aligned}$$

# Online Learning with Multiple Subsequences

Patients arrive one at a time; each belongs to several groups



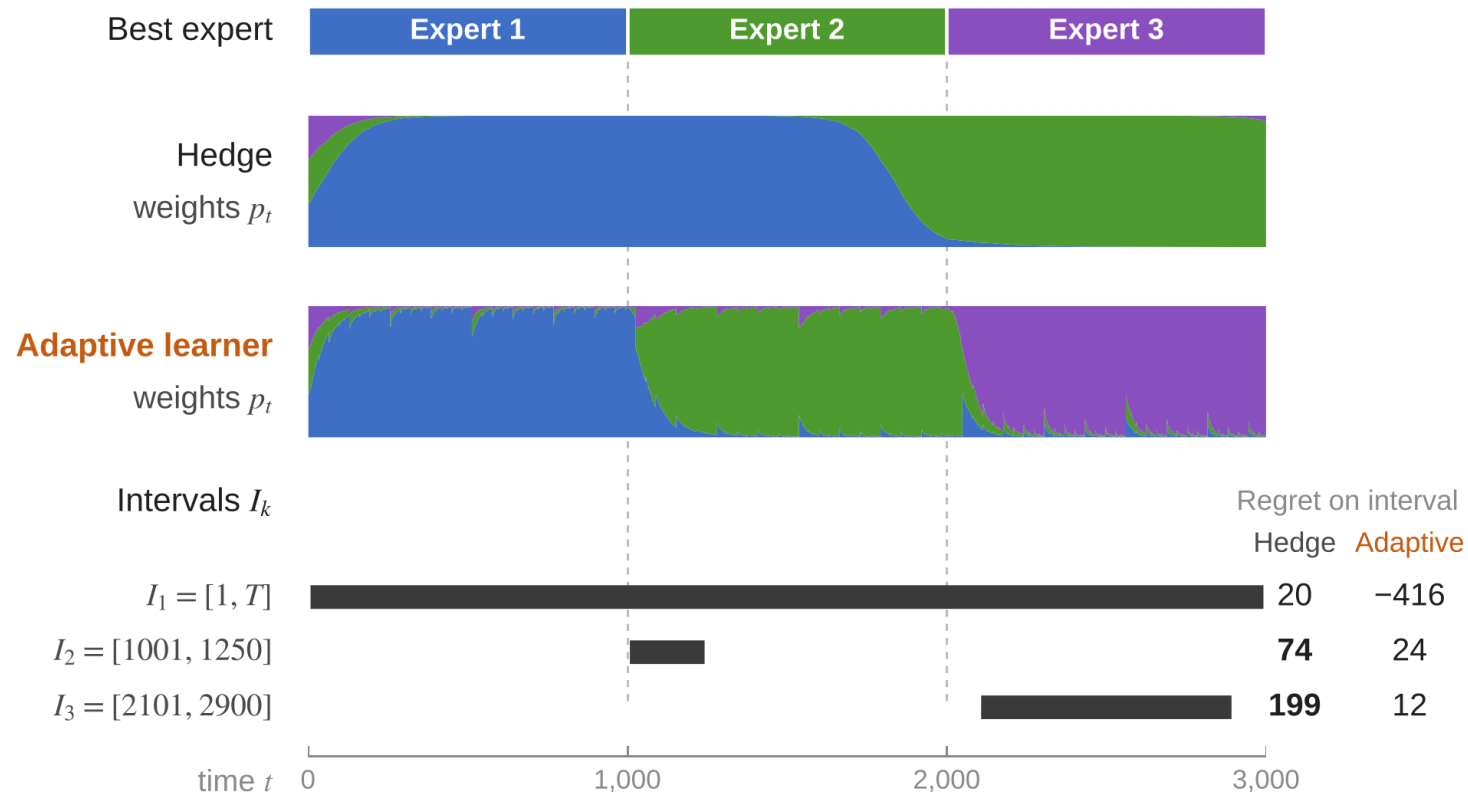
**Why care about group-wise regret?**

# Online Learning with Multiple Subsequences

(a) The best expert changes over time; every interval is a group

$$\text{Adaptive regret} = \max_{I=[s,e] \subseteq [T]} \left( \sum_{t \in I} \langle p_t, \ell_t \rangle - \min_{i \in [N]} \sum_{t \in I} \ell_t(i) \right)$$

= group-wise regret with one group per interval:  $C_t(k) = \mathbf{1}\{t \in I_k\}$



# Is the standard no-regret guarantee sufficient?

No.

Consider loss sequence  $l_t = (0, 1)$  for  $t \leq \frac{T}{2}$   
 $l_t = (1, 0)$  for  $t \in [\frac{T}{2}, T]$

Since  $L_t(1) \leq L_t(2)$  on any  $t$ , if we run exponential weights on it, we have  $p_t(1) \geq p_t(2)$  on any  $t$

This makes  $\sum_{t=\frac{T}{2}}^T \langle p_t, l_t \rangle - \sum_{t=\frac{T}{2}}^T l_t(2) = \Omega(T)$

# Our goal

$$\max_{k \in [K]} \max_{i \in [N]} \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(i)) \leq o\left(\sqrt{T \log(NK)}\right)$$

In fact, the following stronger guarantee is possible:

$$\forall k \in [K] \quad \max_{i \in [N]} \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(i)) \leq o\left(\sqrt{\left(\sum_{t=1}^T C_t(k)\right) \log(NK)}\right)$$

# Recall: No-Regret Best-Response Dynamics

With a known  $G \in \mathbb{R}^{M \times N}$

## p-player (minimizer)

For  $t = 1, 2, \dots$

Choose  $p_t = \operatorname{argmin}_{p \in \Delta_M} p^\top G q_t$

## q-player (maximizer)

For  $t = 1, 2, \dots$

Generate  $q_t \in \Delta_N$  from no-regret alg

Feed no-regret alg with reward  $\underline{G^\top p_t} \in \mathbb{R}^N$

Boosting: Model player  
LP: Variable player

Sample player  
Constraint player

# Recall: No-Regret Best-Response Dynamics

Guarantee:

$$\max_q \sum_{t=1}^T p_t^\top G q \leq T \cdot \text{GameValue}(G) + O(\sqrt{T \log N})$$

//  $\min_p \max_q p^\top G q$

By the regret guarantee of the  $q$ -player,

$$\max_q \sum_{t=1}^T p_t^\top G q - \sum_{t=1}^T p_t^\top G q_t \leq O(\sqrt{T \log N})$$

$$p_t^\top G q_t = \min_p p^\top G q_t \leq \min_p \max_q p^\top G q$$

# Modified Scenario

With a **time-varying** and **unknown**  $G_t \in \mathbb{R}^{M \times N}$

Additional assumption:

$$\forall q \exists p \forall G_t \quad p^\top G_t q \leq V^*$$

**p-player (minimizer)**

For  $t = 1, 2, \dots$

~~Choose  $p_t = \operatorname{argmin}_{p \in \Delta_M} p^\top G_t q_t$~~

Choose  $p_t$  such that  $p_t^\top G_t q_t \leq V^*$

**q-player (maximizer)**

For  $t = 1, 2, \dots$

Generate  $q_t \in \Delta_N$  from no-regret alg

Feed no-regret alg with reward  $G_t^\top p_t \in \mathbb{R}^N$

$$\forall q \min_p \max_{G_t} p^\top G_t q \leq V^*$$

No-regret robust-response dynamics

# Modified Scenario

Guarantee:

$$\max_q \sum_{t=1}^T p_t^\top G_t q \leq TV^* + O(\sqrt{T \log N})$$

By the regret guarantee of the  $\hat{g}$ -player,

$$\max_{\hat{g}} \sum_{t=1}^T p_t^\top G_t \hat{g} - \underbrace{\sum_{t=1}^T p_t^\top G_t \hat{g}_t}_{\leq V^*} \leq O(\sqrt{T \log N})$$

# A “trivial” instantiation of the dynamics

$$G_t \in \mathbb{R}^{N \times N}$$

$$G_t(i, j) = \ell_t(i) - \ell_t(j)$$

**Verify:**  $\forall q \exists p \forall G_t \quad p^\top G_t q \leq 0 = V^*$

$$\begin{aligned} p^\top G_t q &= \sum_{i,j} p(i) q(j) G_t(i,j) \\ &= \sum_{i,j} p(i) q(j) (\ell_t(i) - \ell_t(j)) \\ &= \underbrace{\langle p, \ell_t \rangle - \langle q, \ell_t \rangle}_{\leq 0} \end{aligned}$$

Given  $q$ , just set  $p = q$ .

Guarantee:

$$\max_q \sum_{t=1}^T p_t^\top G_t q \leq O(\sqrt{T \log N})$$

$$\Rightarrow \max_q \sum_{t=1}^T \sum_{i,j} p_t(i) q(j) (\ell_t(i) - \ell_t(j)) \leq O(\sqrt{T \log N})$$

$$\Rightarrow \max_q \sum_{t=1}^T (\langle p_t, \ell_t \rangle - \langle q, \ell_t \rangle) \leq O(\sqrt{T \log N})$$

$p$ -player's regret over loss  $\ell_t$

# A “trivial” instantiation of the dynamics

$$G_t \in \mathbb{R}^{N \times N}$$

$$G_t(i, j) = \ell_t(i) - \ell_t(j)$$

The interaction just looks like below (somewhat trivial):

**p-player (minimizer)**

For  $t = 1, 2, \dots$

Choose  $p_t = q_t$

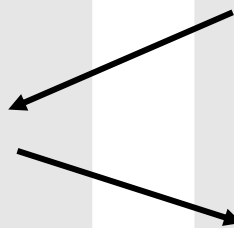
**q-player (maximizer)**

For  $t = 1, 2, \dots$

Generate  $q_t \in \Delta_N$  from no-regret alg

Feed no-regret alg with reward

$$r_t(j) = \langle p_t, \ell_t \rangle - \ell_t(j)$$



# Back to “Online Learning with Multiple Subsequences”

For  $t = 1, 2, \dots, T$ :

Learner observes a context  $C_t \in \{0,1\}^K$

Learner generates a distribution  $p_t \in \Delta_N$

Adversary decides the loss vector  $\ell_t = (\ell_t(1), \dots, \ell_t(N)) \in \mathbb{R}^N$

Adversary reveals  $\ell_t$  to the learner

Learner suffers loss  $\langle p_t, \ell_t \rangle$

$$\begin{aligned} \text{Group-wise Regret} &= \max_{k \in [K]} \left( \sum_{t: C_t(k)=1} \langle p_t, \ell_t \rangle - \min_{i \in [N]} \sum_{t: C_t(k)=1} \ell_t(i) \right) \\ &= \max_{k \in [K]} \max_{i \in [N]} \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(i)) \end{aligned}$$

# Back to “Online Learning with Multiple Subsequences”

$$G_t \in \mathbb{R}^{N \times NK}$$

$$G_t(i, (j, k)) = C_t(k)(\ell_t(i) - \ell_t(j)) \quad i \in [N], j \in [N], k \in [K] \text{ (group)}$$

**Verify:**  $\forall q, C_t \exists p \forall \ell_t \quad \underline{p^\top G_t q} \leq 0$

$$W_t(j) \triangleq \sum_k f(j, k) C_t(k)$$

$$p^\top G_t f = \sum_{i, j, k} p(i) f(j, k) G_t(i, (j, k))$$

$$= \sum_{i, j, k} p(i) f(j, k) C_t(k) (\ell_t(i) - \ell_t(j))$$

$$= \sum_i p(i) \ell_t(i) \sum_{j, k} f(j, k) C_t(k) - \sum_i p(i) \sum_j \ell_t(j) \sum_k f(j, k) C_t(k)$$

$$\approx \sum_i p(i) \ell_t(i) \sum_j W_t(j) - \sum_i p(i) \sum_j \ell_t(j) W_t(j)$$

$$\begin{aligned}
& \sum_i p(i) l_t(i) \sum_j w_t(j) - \sum_i p(i) \sum_j l_t(j) w_t(j) \\
& \quad \text{reindex } i \leftrightarrow j \\
& = \sum_{i,j} p(i) l_t(i) w_t(j) - \sum_{i,j} p(j) l_t(j) w_t(i) \\
& = \sum_{i,j} \left( p(i) w_t(j) - p(j) w_t(i) \right) l_t(i) = 0 \\
& \quad \text{pick } \boxed{p(i) = c \cdot w_t(i)} \\
& \quad \uparrow \text{depend } g \cdot C_t \\
& \quad c w_t(i) w_t(j) - c w_t(j) w_t(i) = 0
\end{aligned}$$

# Back to “Online Learning with Multiple Subsequences”

$$G_t \in \mathbb{R}^{N \times NK}$$

$$G_t(i, (j, k)) = C_t(k)(\ell_t(i) - \ell_t(j))$$

Guarantee:

$$\max_{\mathbf{g}} \sum_{t=1}^T p_t^\top G_t \mathbf{g} \leq O(\sqrt{T \log NK})$$

$$\Rightarrow \max_{\mathbf{g}} \sum_{t=1}^T \sum_{i, j, k} p_t(i) g(j, k) C_t(k) (\ell_t(i) - \ell_t(j)) \leq O(\sqrt{T \log(NK)})$$

$$\Rightarrow \max_{j, k} \sum_{t=1}^T \sum_i p_t(i) C_t(k) (\ell_t(i) - \ell_t(j)) \leq O(\sqrt{T \log(NK)})$$

$$= \max_k \max_j \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(j)) \leftarrow \text{group-wise regret}$$

# Resulting Algorithm

$$p_t(i) \propto \sum_{K: \text{active in } t} \delta_t(i, k)$$

For  $t = 1, 2, \dots$

Choose  $p_t \in \Delta_N$  as

$$p_t(i) = \frac{\sum_k q_t(i, k) C_t(k)}{\sum_{j, k} q_t(j, k) C_t(k)} = \frac{w_t(i)}{\sum_j w_t(j)}$$

For  $t = 1, 2, \dots$

Generate  $q_t \in \Delta_{NK}$  from no-regret alg

Feed no-regret alg with reward  
 $r_t(j, k) = C_t(k)(\langle p_t, \ell_t \rangle - \ell_t(j))$

# A No-Groupwise-Regret Algorithm

Initialize  $q_t(i, k) = 1$  for all  $(i, k) \in [N] \times [K]$

For  $t = 1, 2, \dots, T$ :

Learner observes  $C_t \in \{0, 1\}^K$

Learner plays distribution

$$p_t(i) \propto \sum_{k=1}^K C_t(k) q_t(i, k)$$

Adversarial reveals  $\ell_t \in \mathbb{R}^N$

Update  $q_{t+1}(i, k) = q_t(i, k) \exp \left( \eta C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(j)) \right)$

$$\max_{k \in [K]} \max_{i \in [N]} \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(i)) \leq O \left( \sqrt{T \log(NK)} \right) \quad \text{with optimal choice of } \eta$$

# How to Achieve A More Adaptive Performance?

$$\forall k \in [K] \quad \max_{i \in [N]} \sum_{t=1}^T C_t(k) (\langle p_t, \ell_t \rangle - \ell_t(i)) \leq O \left( \sqrt{\left( \sum_{t=1}^T C_t(k) \right) \log(NK)} \right)$$

Recall exponent weights guarantees

$$p_{t+1}(i) \propto p_t(i) \exp(-\gamma \ell_t(i)) \rightarrow \sum_t \langle p_t, \ell_t \rangle - \sum_t \ell_t(i^*) \leq \frac{\log N}{\gamma} + 2 \sum_t \sum_i p_t(i) \ell_t(i)^2 \quad \hookrightarrow i^*$$

$$\rightarrow \sum_t \langle p_t, \ell_t \rangle - \sum_t \ell_t(i^*) \leq \frac{\log N}{\gamma} + \gamma \sum_t \ell_t(i^*)^2 \quad \hookrightarrow i^*$$

$$p_{t+1}(i) \propto p_t(i) \exp(-\gamma \ell_t(i) - \gamma^2 \ell_t(i)^2)$$

# How to Achieve A More Adaptive Performance?

$$\star \quad \sum_t \langle p_t, l_t \rangle - \sum_t l_t(i^*) \leq \sqrt{\left( \sum_t l_t(i^*)^2 \right) \log(NT)} \quad \forall i^*$$

If we use this in the update of  $\hat{\theta}_t$ , then

$$\begin{aligned} \forall k, i \quad \sum_t C_t(k) \left( \langle p_t, l_t \rangle - l_t(\underline{i}) \right) &\leq \sqrt{\sum_t \left( C_t(k) \left( \langle p_t, l_t \rangle - l_t(\underline{i}) \right) \right)^2 \log(NTk)} \\ &\leq \sqrt{\left( \sum_t C_t(k) \right) \log(NTk)} \end{aligned}$$

Blum and Lykouris. [Advancing Subgroup Fairness via Sleeping Experts](#). 2020.

 Chen, Luo, Wei. [Impossible Tuning Made Possible: A New Expert Algorithm and Its Applications](#). 2021.

# Application: Interval Regret (a.k.a. Adaptive Regret)



$$\sum_{t=t_1}^{t_2} \langle p_t, \ell_t \rangle - \min_{i \in [N]} \sum_{t=t_1}^{t_2} \ell_t(i) \lesssim O(\sqrt{t_2 - t_1}) \quad \text{for all } 1 \leq t_1 < t_2 \leq T ?$$

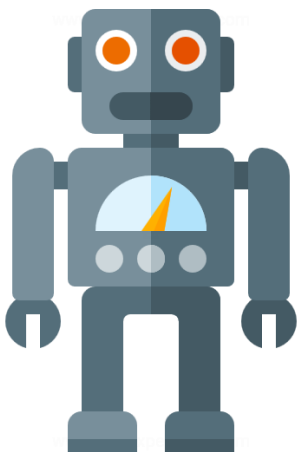
Let every group correspond to an interval  $[t_1, t_2]$  (there'll be  $\frac{T(T-1)}{2}$  groups)

$$\sum_t \langle p_t, \ell_t \rangle - \sum_t \min_i \ell_t(i) \lesssim \sqrt{T_1} + \dots + \sqrt{T_S} \lesssim \sqrt{ST}$$



# of times the best expert changes

$$\log \left( NT \cdot \frac{T(T-1)}{2} \right) = O(\log NT)$$



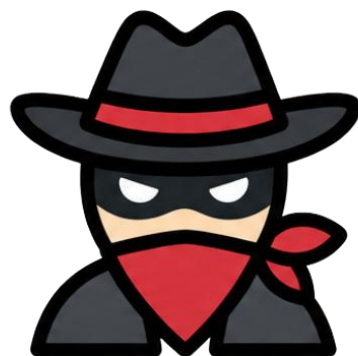
Learner



Auditor

monitors if  $\exists k, i$  such that

$$\sum_t \left[ c_t(k) \left( \langle p_{t, l_t} \rangle - l_t(i) \right) \right] \text{ is large}$$



Adversary

$q_t$

# Swap Regret

# Internal Regret / Swap Regret

For  $t = 1, 2, \dots, T$ :

Learner generates a distribution  $p_t \in \Delta_N$

Learner samples an action  $i_t \sim p_t$

Adversary decides the loss vector  $\ell_t = (\ell_t(1), \dots, \ell_t(N)) \in \mathbb{R}^N$

Adversary reveals  $\ell_t$  to the learner

$$\begin{aligned} \text{Internal Regret for Action } i &= \sum_{t: i_t=i} \ell_t(i) - \min_{j \in [N]} \sum_{t: i_t=i} \ell_t(j) \\ &= \max_{j \in [N]} \sum_{t=1}^T C_t(i)(\ell_t(i) - \ell_t(j)) \end{aligned}$$

= Group-wise regret where “Group  $i$ ” consists of rounds in which the learner selects action  $i$

# Internal Regret / Swap Regret

For  $t = 1, 2, \dots, T$ :

Learner generates a distribution  $p_t \in \Delta_N$

Learner samples an action  $i_t \sim p_t$

Adversary decides the loss vector  $\ell_t = (\ell_t(1), \dots, \ell_t(N)) \in \mathbb{R}^N$

Adversary reveals  $\ell_t$  to the learner

$$\begin{aligned} \text{Swap Regret} &= \sum_{i=1}^N [\text{Internal Regret for Action } i] \\ &= \sum_{t=1}^T \ell_t(i_t) - \min_{\phi: [N] \rightarrow [N]} \sum_{t=1}^T \ell_t(\phi(i_t)) \end{aligned}$$

**Using our previous algorithm for group-wise regret (?)**