You against the world

Route-picking

betting on stock market

Repeatedly make decisions, experience consequences

Adaptive decision making

\[ \text{day}^t \xrightarrow{\text{choose among}} \text{set of possible actions} \]

\[ \text{find out} \xrightarrow{\text{your cost/loss for day}} \text{(possibly loss for all alternatives)} \]

How well can you do?

Formally, each day have \( n \) possible actions to choose from

\[ \text{choose } i^t \in [n] \]

\[ \text{learn } l^t = (l_1^t, \ldots, l_n^t) \]

\[ \text{Loss} = \sum_{t=1}^{T} l_{i_t}^t \]

Benchmarks:

\[ \text{OPT}_d = \sum_{t=1}^{T} \min_{i} l_i^t \]

\[ L^* = \min_{i} \sum_{t=1}^{T} l_i^t \]

We'll assume each \( l_i^t \in [0,1] \)
Consider max algo \[ \min \{ L_i^+ \} \] performance of algo only

Can't compete with \( \text{OPT}_d \) even with \( n = 2 \)

\[ L_i^+ \text{ equally likely to be 0 or 1} \]
\[ L_i^+ = 1 - L_i \]

Alg has loss \( \frac{I}{2} \) \( \text{OPT}_d = 0 \)

Instead consider regret

\[ \text{Regret} = \text{Regret}(A, \{L_i^+\}) = \frac{I}{2} \text{ Loss of } A \]
\[ \text{Regret} = \min_i L_i^+ \]

Multiplicative Weight Updates Alg

Fix \( \varepsilon < \frac{1}{2} \); \( n \) possible actions. On each day \( t \), \( w_i^t \) = weight of expert \( i \)

Initially \( w_i^0 = 1 \) \( \forall i \)

On day \( t \), use mixed strategy \( \tilde{p}_i^t \) where
\[
\tilde{p}_i^t = \frac{w_i^{t+1}}{\sum w_k^{t+1}}
\]

For each action \( i \), observe the loss \( L_i^t \in [0, I] \)

And update the weight as follows:
\[
w_i^t := w_i^{t+1} \cdot e^{-\varepsilon L_i^t}
\]
Theorem: Given any loss sequence \( \{l^t\}_{t=1}^T \), let \( L^T = \sum_{t=1}^T p^t l^t \)

where \( p^t \) are the prob distr used by MWU

Then \( \forall i \):

\[
L^T \leq l^T_i + \frac{T \varepsilon}{8} + \frac{\ln n}{\varepsilon}
\]

where \( l^T_i = \sum_{t=1}^T l^t_i \)

\( \frac{T \varepsilon}{8} \) will prove

\( \frac{T \varepsilon}{2} \)

Corollary: with \( \varepsilon = \sqrt{\frac{8 \ln n}{T}} \), \( \forall \{l^t\}_{t=1}^T \)

\[
\text{Regret (MWU)} \leq \sqrt{\frac{T \ln n}{2}}
\]

Proof of Thm:

Let \( W^t = \sum_{i=1}^n w^t_i = \sum_{i=1}^n w^{t+1}_i e^{-\varepsilon l^t_i} \)

\[
\frac{W^t_{t+1}}{W^t_{t+1}} = \sum_{i=1}^n w^{t+1}_i = \frac{\sum_{i=1}^n p^t_i e^{-\varepsilon l^t_i}}{E(e^{-\varepsilon X^t})} = e^{-\varepsilon p^t X^t} E(e^{-\varepsilon (X^t - p^t X^t)})
\]

where \( X^t = l^t_i \) with prob \( p^t_i \)

\[
\leq e^{-\varepsilon p^t X^t} e^{\frac{\varepsilon^2}{8}}
\]

[actually \( \leq e^{\frac{\varepsilon^2}{8}} \)]
\[ \Rightarrow W^T e \leq \mathbb{E} W_{1}^{T-1} \]

\[ \leq e \mathbb{E} e^T \mathbb{E} . n \]

OTOH \quad W^T w_i = e^{-3L_i^T}

So \quad e \leq e \mathbb{E} e^T \mathbb{E} n

Taking logs \quad -3L_i^T \leq -3L + \frac{T e^2}{\alpha} + \frac{\ln n}{\varepsilon}

\[ \Rightarrow L^T \leq L_i^T + \frac{T e^2}{\alpha} + \frac{\ln n}{\varepsilon} \]

\[ \frac{T e^2}{\alpha} = \frac{\ln n}{\varepsilon} \Rightarrow \varepsilon^2 = \frac{\alpha}{T} \ln n \equiv \varepsilon = \sqrt{\frac{\alpha \ln n}{T}} \]

\[ \Rightarrow L^T \leq L_i^T + \sqrt{\alpha T \ln n} \]

**Hoeffding Lemma:** Let \( X \) be r.v. with \( E(X) = 0 \) and \( |X| \leq 1 \)

Then \( E(e^{\lambda X}) \leq e^{\lambda^2 / 2} \)

Proof: \( e^{\lambda X} \) is convex \( \Rightarrow e^{\lambda X} \leq \frac{(1+x) e^{\lambda x} + (1-x) e^{-\lambda x}}{2} \) for \( x \in [-1,1] \)

\[ e^{\lambda X} \leq \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \]

Since \( |X| \leq 1 \) & \( E(X) = 0 \)

\[ E \left[ \frac{(1+x) e^{\lambda x} + (1-x) e^{-\lambda x}}{2} \right] = e^{\lambda \frac{\lambda^2}{2}} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \]

\[ \leq \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \left( \frac{e^{\lambda^2 / 2}}{2} \right) \leq e^{\lambda^2 / 2} \]

\[ (2k)! > 2^k \cdot k! \]
$V^T$: unavoidable

Say $n = 2$: flip coin each day

Any easy hae exp loss $\frac{T}{2}$

$E[\min(\#H's, \#T's)] = \frac{T}{2} - c\sqrt{T}$

In fact above bound $\sqrt{\frac{T\log}{2}}$ is tight up to lower order terms

(analyse case where each $f_i^+$ indepuniform $\{0,1\}$)

Applications: (started with game playing)

1) Route-picking

2) Learning: choosing prediction from class $\mathcal{F}$ from $\{f_1, ..., f_k\}$

Given $x_t$ (data for day $t$), choose $f_t$ in $\mathcal{F}$, incur loss $l(f_t(x), y_t)$

low regret means that in hindsight do nearly as well as best $f_t$ in class

3) Pricing: sequence of customers, each with value $v_t$ for item you're selling
Application to setting with gains \( \varepsilon \) where \( \forall t \quad \max_k \ g_i^t - \min_j g_j^t \leq \varepsilon \\

Let \( g_{\text{max}}^t = \max_k g_k^t \)

Define \( \ell_i^t = \frac{1}{\varepsilon} (g_{\text{max}}^t - g_i^t) \)

\[
\text{Regret} (\text{MW}) \leq \sqrt{\frac{T \log T}{2}} \\
\Rightarrow L_{\text{MW}}^T \leq L_j^T + \sqrt{\frac{T \log T}{2}} \]

\[
\sum_t (g_{\text{max}}^t - g_j^t) \leq \sum_t (g_{\text{max}}^t - g_j^t) + \varepsilon \sqrt{\frac{T \log T}{2}} \\
\Rightarrow \text{MW gain} \geq G_j^T - \varepsilon \sqrt{\frac{T \log T}{2}} \quad \forall j
\]

Application: portfolio selection

- investing money in stocks/commodities/currencies
- each day reinvest
\[
\frac{\text{closing price}}{\text{opening price}} = \text{proportion}
\]

<table>
<thead>
<tr>
<th>Stock 1</th>
<th>1.5</th>
<th>(\frac{1}{2})</th>
<th>at end of day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock 2</td>
<td>1</td>
<td>0</td>
<td>$1 \Rightarrow \frac{1}{2} \cdot 1.5 + \frac{1}{2} \cdot 1 = $1.25</td>
</tr>
<tr>
<td>Stock 3</td>
<td>1</td>
<td>(\frac{1}{2})</td>
<td></td>
</tr>
<tr>
<td>Stock 4</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

\[
p_i^+ = \text{fraction of wealth you put into stock } i \text{ on day } t
\]

\[
r_i^+ = \frac{\text{closing price of stock } i \text{ on day } t}{\text{opening price on day } t}
\]

If invested all your money in \(i\),
\[
\frac{\text{final wealth}}{\text{initial}} = r_1^+ r_2^+ \ldots r_t^+
\]

\[
\log \frac{W_t}{W_0} = \sum_{t} \log r_t^+
\]

If we choose portfolio each day using MW,
\[
\sum_{t=1}^{T} \sum_i p_i^+ \log r_t^+ \geq \sum_t \log r_t^+ - \varepsilon \quad \forall i
\]

\[
\leq \log \left( \sum_i p_i^+ r_t^+ \right)
\]

\[
\log \text{concave function}
\]
\[
= \sum_{+i} \log \left( \frac{p_i r_i}{r_i} \right) = \log \prod_{+i} r_i - 3
\]

\[
= \log \prod_{i \neq +i} r_i
\]

\[
\frac{\text{Final weight}}{w_0} \geq \frac{w_i}{w_0} e^{-3}
\]

---

Application to playing 2-player zero-sum games

Game defined by matrix

Row player is \( R \), col player is \( C \); entry is payoff to \( R \)
Example 1: Penalty Kicks

![Penalty Kicks Diagram]

Example 2: Hide & Seek

- Cop picks street or avenue
- Robber picks a safehouse
- Cop's payoff is 1 if picks street or avenue where robber is hiding

König's Lemma: max matching = minimum line (vertex) cover

Pf: max matching ≤ min vertex cover

if no matching of size r in purple part, can replace some rows in r' by columns in purple
Example 3 “Smuggler vs border guard” (Blum)

Graph $G$, sources, sink $t$

smuggler chooses path

border guard chooses edge to watch

If edge is in path, guard wins, else smuggler wins

Menger’s Thm: Min $s$-$t$ cut (edge) = max # edge disjoint $s$-$t$ paths

minimax opt strategy for row player

$V_R = \max_p \min_j \sum_i p_{aij}$

What $R$ would play if she had to reveal strategy first

minimax opt strategy for col player

$V_C = \min_q \max_i \sum_j a_{ij} q_j$

What $C$ would play if he had to reveal strategy first

Clearly $1 < 2$
von Neumann 1928  \[ V_R = V_C = V \]

no loss to publish strategy

**Proof:** Suppose \( V_C > V_R + \varepsilon \) Scale payoffs to \([0,1]\)

Have C play RWM against R responding optimally

In \( T \) steps: \( C \)'s loss \( \leq \sum_{t=1}^{T} a_{i:t} + c\sqrt{T \ln n} \quad \forall j \)

\( \leq TV_R + \sqrt{T \ln n} \)

\( C \)'s loss = \( R \)'s gain \( \geq TV_C \)

\( \Rightarrow \) \( TV_C \leq TV_R + \sqrt{T \ln n} \)

\( V_C - V_R < \sqrt{\frac{\ln n}{T}} \)

\( V_C - V_R \leq \sqrt{\frac{\ln n}{T}} \quad \rightarrow \quad \text{for } T \text{ sufficiently large} \)

Note: if opponent not playing optimally, do much better than \( V \)

MWU gives fast way to compute approx optimal strategies

Let \( q^* = \frac{1}{T} \sum_{t=1}^{T} p_t \) and let \( q^* \) be best response of R to \( q^* \)

\[ \left( \frac{1}{T} \sum_{t=1}^{T} p_t \right) a_{i:0}^* \leq \frac{1}{T} \sum_{j=1}^{n} p_j a_{i:j}^* \leq V + \varepsilon \]

\( i^* \) is best response at each step \( \leq \frac{\varepsilon}{T} \sum_{j=1}^{n} a_{i:j} + \varepsilon \)
Let \( \mathcal{P}^* = \{ f : |f_i = i^3 \} \) be empirical distribution.

\[
\forall_j \frac{1}{T} \sum_{t=1}^{T} a_{i,j} > \frac{1}{T} \sum_{t=1}^{T} I_{p_j \in \mathcal{P}^*} a_{i,j} - \epsilon
\]

\[
\geq V - \epsilon
\]

Solving LP

\[
\exists x \succeq 0 \text{ s.t. } \underbrace{A x \succeq b}_{\text{m.m.}}
\]

Use MWU to find \( x \succeq 0 \) s.t. \( \forall i \ a_i x \succeq b_i - \epsilon \) or prove system infeasible.

Diagram:

- Requires \( \overrightarrow{p} \in \mathbb{R}^m \)
- Oracle:
  - Yes: \( \exists x \succeq 0 \text{ s.t. } p^T A x \succeq p^T b \) and \( |a_i x - b_i| \leq \epsilon \)
  - No system infeasible
Thm: Assume \( 0 < \eta \leq 2 \gamma \)

Then Algorithm either finds \( v_{opt} \) s.t. \( \forall i \ a; x > b \), \( \varepsilon \)

or correctly concludes that system is infeasible

\[ O\left( g^2 \frac{\log m}{\varepsilon^2} \right) \] calls to oracle with additional processing time \( O(m) \) per call

Call Oracle
Get \( x^{(1)} \)

MWU

\[ \frac{p^{(t)}}{dist \mbox{ even}} \rightarrow \]

\[ \frac{1}{i} [a; x^{(t)} - b;] \]

\[ p^{(t)} l^{(t)} = p^{(t)} \cdot \frac{1}{g} [Ax^{(t)} - b] \geq 0 \] guarantee g oracle

Thm about MWU guarantees

\[ 0 \leq \frac{1}{g} \sum_{t=1}^{T} (a; x^{(t)} - b; i) + \eta \frac{1}{g} \sum _{t=1}^{T} |a; x^{(t)} - b;| + \ln m \frac{m}{g} \]

\[ \leq \sum_{t=1}^{T} (a; x^{(t)} - b; i) + \frac{\eta g T}{g} + \ln m \frac{m}{g} \]

\[ \Rightarrow \frac{1}{T} \sum_{t=1}^{T} (a; x^{(t)} - b; i) \geq \eta g + \frac{p \ln m}{g} \Rightarrow a; \left( \frac{1}{T} \sum_{t=1}^{T} x^{(t)} \right) - b; \geq \eta g + \frac{p \ln m}{g} \]
choose $\eta = \frac{\varepsilon}{4p}$ and $T = \frac{8p^2 \ln m}{\varepsilon^2}$

and rhs becomes $\frac{\varepsilon}{4} + \frac{p \ln m (16p^2 - p^2)}{\eta^2 8p^2 ln m} = \frac{\varepsilon}{4} + \frac{2p}{\eta}$
Example: \( \text{flow} \) \( \max \sum_{p} y_{p} \)

\[
\sum_{p \in e} y_{p} \leq c_{e}
\]

\[
-\sum_{p \in e} y_{p} \geq -1
\]
Multi-armed bandit Problem

reduction to standard setting

\[ q^+ \in [0, h] \]

\[ p^+ \]

\[ g^+ \]

\[ \hat{g}^+ \]

\[ \bar{g}^+ \]

\[ \hat{g} \]

\[ E(\hat{g}^+) = g^+ \]

\[ \sum_{i} \hat{q}_i^+ \geq E(\max_{i} \hat{g}_i^+) - \frac{n \log T}{\delta} \]

\[ \sum_{i} \hat{q}_i^+ \geq (1-\delta^*) \sum_{i} \hat{p}_i^+ \hat{g}_i^+ = (1-\delta^*) \sum_{i} \hat{p}_i^+ E(\hat{g}_i^+) \]
Claim: \( E(\mathbf{G}_j^T) = \mathbf{G}_j^T \)  
\( E(\mathbf{G}_j^0) = (1-q_j^0) \cdot 0 + q_j^0 \left( \frac{q_j^0}{q_j^0 + \bar{q}_j^0} \right) = q_j^0 \).

\[
E[\max \mathbf{G}_j^T] \geq \max_j E(\mathbf{G}_j^T) = \max_j \mathbf{G}_j^T
\]