ECE 493 Cheatsheet

Probability

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Agent preferences

Lottery

Let O be a set of outcomes. Lottery A is a probability distribution over outcomes.

- $A \prec B$ if agent strictly prefers A over B
- $A \preceq B$ if agent weakly prefers A over B
- $A \sim B$ if agent is indifferent between A and B

VNM Rationality

- Completeness: $A \prec B$ or $A \preceq B$ or $A \sim B$
- Transitivity: $A \prec B$ and $B \prec C \Longrightarrow A \prec C$
- Independence: $A \prec B$ iff $pA + (1-p)C \prec pB + (1-p)C$
- Continuity: $A \prec B \prec C \Longrightarrow \exists p \in [0,1], B \sim pA + (1-p)C$
- Betweenness: $p \in (0,1) \Longrightarrow A \prec pA + (1-p)B \prec B$
- Monotonicity: $p > q \Longrightarrow pA + (1-p)B \preceq qA + (1-q)B$

VNM utility theorem

For any agent there exists utility function u s.t.

- $u(A) = u(\sum p_k o_k) = \sum_{p_k} u(o_k)$
- $u(A) \geq u(B)$ iff $A \preceq B$

Risk attitude

Let $A = px + (1-p)y$, $z = \mathbb{E}(A)$

- Risk neutral: $u(A) = u(z)$
- Risk averse: $u(A) < u(z)$
- Risk seeking: $u(A) > u(z)$

Normal-form games

- Consists of N agents
- Agent i has available actions A_i
- Outcome is an action profile A of all agents
- Game is zero-sum if $\sum_i u_i(A) = 0 \ \forall A$

Dominant strategy equilibrium

- s_i strictly dominates s_j if $u_{i(s_i)} > u_{i(s_j)}$ for all j
- s_i weakly dominates s_j if $u_{i(s_i)} \geq u_{i(s_j)}$ for all j and $u_{i(s_i)} > u_{i(s_j)}$ for at least one j
- s_i is a strict/weak dominant strategy if it strictly/weakly dominates all other strategies
- s is a **dominant strategy equilibrium** if s_i is the dominant strategy for all i
- s if exists can be found by iterated elimination of dominated strategies
- Order of elimination does not matter for strictly dominant strategies

Nash equilibrium

- s_i^* is a **best response** to s_{-i} if $u_i(s_i^*, s_{-i}) \geq u_i(s_i, s_{-i})$ for all $s_i \in A_i$
- Nash equilibrium: s^* such that $u_i(s^*, s_{-i}^*) \geq u_i(s_i', s_{-i}^*) \ \forall i, s_{i'}$
  - Nash theorem: every finite game has a Nash equilibrium
- To compute a Nash equilibrium,
  1. Assign probabilities p_i for each action for agent i
  2. Equate utilties of other agent given each action and their expected utilities
- Strict NE: $u_i(s^*, s_{-i}^*) > u_i(s_i', s_{-i}^*) \ \forall i, s_i' \neq s_i^*$
  - Must be pure strategy NE
- Strong NE: no group of agents win by unilateral deviation
- Stable NE: no agent win by any small deviations of any agent

Price of anarchy

- Braess's paradox: adding new zero-cost links to a network can increase travel time
- Price of anarchy: ratio between worse NE performance and optimal performance

Minmax and maxmin strategies

- Maxmin strategy: $\operatorname{argmax}_{s_i} \min_{s_{-i}} u_i(s_i, s_{-i})$
- Minmax strategy: $\operatorname{argmin}_{s_{-i}} \max_{s_i} u_i(s_i, s_{-i})$
- Minimax theorem: in any NE of any finite two-player zero-sum game, minmax value and maxmin value are equal

Rationalizability

- Rationalizable strategy: best response to some belief about strategies of other agents

Correlated equilibrium

- Recommendations: distribution $\pi \sim \Delta(A)$ over strategy profiles
  - Given r sampled from $R \in \pi$, each agent forms belief of others' strategies by

$$\pi(r_{-i} \mid r_i) = \frac{\pi(r_i, r_{-i})}{\sum_{r'_{-i} \in A_{-i}} \pi(r_i, r'_{-i})}$$

- Correlated equilibrium: $\forall i, r_i, r_i'$,

$$\sum_{r_{-i} \in A_{-i}} \pi(r_{-i} \mid r_i)[u_i(r_i, r_{-i}) - u_i(r_i', r_{-i})] \geq 0$$

  - π is CE if $\mathbb{E}_{a \sim \pi}[u_i(A)] \geq \mathbb{E}_{a \sim \pi}[u_i(a_i', a_{-i}) \mid a_i]$
  - No agent can benefit from deviating from recommendation assuming other agents follow their recommendations
- Coarse correlated equilibrium: π such that $\mathbb{E}_{a \sim \pi}[u_i(A)] \geq \mathbb{E}_{a \sim \pi}[u_i(a_i', a_{-i})]$
  - No agent can benefit from not receiving recommendation at all assuming other agents follow their recommendations

MILP for games

Weak dominance

$$\begin{aligned}
\max \quad & \sum_{a_{-i} \in A_{-i}} \left[\left(\sum_{a_i \in A_i} p_{a_i} u_i(a_i, a_{-i}) \right) - u_i(s_i, a_{-i}) \right] \\
\text{s.t.} \quad & \sum_{a_i \in A_i} p_{a_i} u_i(a_i, a_{-i}) \geq u_i(s_i, a_{-i}) & \forall a_{-i} \in A_{-i} \\
& \sum_{a_i \in A_i} p_{a_i} = 1 \\
& p_{a_i} \geq 0 & \forall a_i \in A_i
\end{aligned}$$

- If optimal solution is strictly positive, then s_i is weakly dominated

Strict dominance

$$\begin{aligned}
\max \quad & \varepsilon \\
\text{s.t.} \quad & \sum_{a_i \in A_i} p_{a_i} u_i(a_i, a_{-i}) \geq u_i(s_i, a_{-i}) + \varepsilon & \forall a_{-i} \in A_{-i} \\
& \sum_{a_i \in A_i} p_{a_i} = 1 \\
& p_{a_i} \geq 0 & \forall a_i \in A_i
\end{aligned}$$

- If optimal solution is strictly positive, then s_i is strictly dominated

Maxmin

For agent i,

$$\begin{aligned}
\max \quad & u_i \\
\text{s.t.} \quad & \sum_{a_i \in A_i} p_{a_i} u_i(a_i, a_{-i}) \geq u_i & \forall a_{-i} \in A_{-i} \\
& \sum_{a_i \in A_i} p_{a_i} = 1 \\
& p_{a_i} \geq 0 & \forall a_i \in A_i
\end{aligned}$$

- Polynomial time
- Can compute NE for two player zero-sum games

Nash equilibrium

$$\begin{aligned}
\max \quad & f(u_1, u_2) \\
\text{s.t.} \quad & p_{a_i} \geq 0 & \forall i, a_i \in A_i \\
& \sum_{a_i \in A_i} p_{a_i} = 1 & \forall i \\
& \sum_{a_{-i} \in A_{-i}} p_{a_{-i}} u_i(a_i, a_{-i}) = u_{a_i} & \forall i, a_i \in A_i \\
& u_{a_i} \leq u_i & \forall i, a_i \in A_i \\
& p_{a_i} \leq b_{a_i} & \forall i, a_i \in A_i \\
& u_i - u_{a_i} \leq M(1 - b_{a_i}) & \forall i, a_i \in A_i \\
& b_{a_i} \in \{0, 1\} & \forall i, a_i \in A_i
\end{aligned}$$

- p_{a_i} is probability of choosing action a_i in mixed strategy
- u_{a_i} is utility of pure strategy a_i
- u_i is utility of the strategy

- b_{a_i} is the mixed strategy support
- M is a large constant

Correlated equilibrium

$$\begin{aligned} \max \quad & f(u_1, ..., u_N) \\ \text{s.t.} \quad & \sum_{a_{-i} \in A_{-i}} p_a(u_i(a) - u_i(t_i, a_{-i})) \geq 0 \quad \forall i, a_i, t_i \in A_i \\ & \sum_{a \in A} p_a = 1 \\ & p_a \geq 0 \quad \forall a \in A \end{aligned}$$

Extensive-form games

- Played sequentially
- Strategies can depend on previous actions
- Every finite extensive-form game has a PSNE

Subgame perfect equilibrium

- Standard NE only requires root node best response
- s^* is SPE if it is NE for every subgame
- Found using backward induction
- Backward induction limitation: tie breaking affects result

Imperfect-info games

- Some nodes are linked together as info sets
  - Players don't know where on the info set they are
- Mixed strategy: randomize then play
- **Behavioral strategy**: randomize actions every info set
  - Expresses different set of strategies compared to mixed strategies
- In games with perfect recall, behavioral strategies are equivalent to mixed strategies

Repeated games

- Agents play a stage game repeatedly
- If stage game has NE, an SPE is repeating the NE strategy
  - Memoryless/stationary strategy

Finite repeated games

- May have discount factor δ
- $u_i = \sum_{r=1}^{R} \delta^{r-1} u_i^{(r)}$

Infinite repeated games

- Future-discounted utility: $u_i = \sum_{r=1}^{\infty} \delta^{r-1} u_i^{(r)}$
- s^* is SPE iff there are no profitable one-shot deviations for each subgame and agent

Strategies

- Grim trigger strategy: punish other player forever if they deviate
  - Done by switching to strategy which reduce utility of other player even more
- Tit-for-tat strategy: mirror opponent's last type of strategy

Folk theorem

- u is **feasible** if there exist rational, non-negative distribution α such that $\forall i, u_i = \sum_{a \in A} \alpha_a u_i(a)$.
  - Convex hull of possible outcomes: $U = \text{Conv}\{u \in \mathbb{R}^N \mid \exists a \in A, u(a) = u\}$
- u is **enforceable** and **individually rational** if $u_i \geq \underline{v}_i \ \forall i$, where $\underline{v}_i$ is the minmax value of agent i
  - Can be enforced using grim trigger
- For infinitely repeated games with average utilities, if u is both feasible and enforceable, then u is the utility profile of some NE
  - May not be SPE

Stochastic games

- State set S and action set A
- Action a at state s leads to s' with probability $p(s, a, s')$
- Each transition has a reward $r(s, a, s')$
- Use future-discounted utility

Policies

- Stationary policy: choose action based on current state
- Value function:

$$\begin{aligned} V^\pi(s) &= \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} r_t(s_t, \pi(s_t), s(t+1)) \mid s_0 = s\right] \\ &= \sum_{s'} p(s, \pi(s), s')(r(s, a, s') + \delta V^\pi(s')) \end{aligned}$$

- State-action value function:

$$Q^\pi(s, a) = \sum_{s'} p(s, a, s')(r(s, a, s') + \delta V^\pi(s'))$$

Policy evaluation

- Find value of policy V^π

$V_0^\pi(s) \leftarrow 0 \ \forall s \in S$
repeat until $V^\pi(s)$ converges $\forall s \in S$
  for $s \in S$
    $V_t^\pi(s') \leftarrow \sum_{s'} p(s, \pi(s), s')(r(s, \pi(s), s') + \delta V_{t-1}^\pi(s'))$

Value iteration

$V_0(s) \leftarrow 0 \ \forall s \in S$
repeat until $V(s)$ converges $\forall s \in S$
  for $s \in S$
    $V_t(s') \leftarrow \max_{a \in A} \sum_{s'} p(s, a, s')(r(s, a, s') + \delta V_{t-1}^\pi(s'))$
for $s \in S$
  $\pi^*(s) \leftarrow \text{argmax}_{a \in A} \sum_{s'} p(s, a, s')(r(s, a, s') + \delta V_{t-1}^\pi(s'))$

Strategies

- $h_t = [s_0, a_0, ...a_{t-1}, s_t] \in H_t$ is the history of the game of t stages
- Set of all deterministic strategies for agent i is $\prod_{t, H_t} A_i$
- Behavioral strategy: $\pi_i(h_t a_i)$ is probability of playing a_i for h_t
- Markov strategy: behavioral strategy that only depends on current state s_t
- Stationary strategy: Markov strategy that is time-independent
- Markov perfect equilibrium: Markov strategy such that it is NE regardless of starting state
- Every n-player, general-sum, discounted-reward game has a MPE

Computing equilibrium

- Easier cases:
  - 2P, general-sum, discounted, single-controller
  - 2P, general-sum, discounted, separable-reward, state-independent transition
  - 2P, zeros-sum, discounted

Shapley algorithm

- Compute MPE for two-player zero-sum games by value iteration

$V_0(s) \leftarrow$ random distribution $\forall s$
repeat until $V(s)$ converges $\forall s$
  for $s \in S$
    Compute matrix game $G(s, V_{t-1})$
    $u(s, a) \leftarrow r(s, a) + \delta \sum_{s'} p(s, a, s') V_{t-1}^\pi(s')$
  for $s \in S$
    $V_t^\pi(s) \leftarrow \max_{\pi_1} \min_{\pi_2} u(s, \pi_1, \pi_2)$ (NE)

Pollatschek & Avi-Itzhak Algorithm

- Compute MPE by policy iteration

$V(s) \leftarrow$ random distribution $\forall s$
repeat until $\pi_1(s), \pi_2(s)$ converges $\forall s$
  for $s \in S$
    Compute matrix game $G(s, V)$
    $\pi_1(s) \leftarrow$ maxmin strategy of agent 1 in $G(s, V)$
    $\pi_2(s) \leftarrow$ minmax strategy of agent 1 in $G(s, V)$
  $V(s) \leftarrow$ policy evaluation for π_1, π_2

Bayesian games

- Agents can have private types θ
- Update belief μ based on other agent types

Equilibria

- Ex-post: agents know everyone's types

$$\mathbb{E}[u_i(s, \theta)] = \sum_{a \in A} \prod_{j \in N} s_j(a_j \mid \theta_j) u_i(a, \theta)$$

- Interim: agents know about own type

$$\mathbb{E}[u_i(s, \theta_i)] = \sum_{\theta_{-i} \in \Theta_{-i}} p(\theta_{-i} \mid \theta_i) \mathbb{E}[u_i(s, (\theta_i, \theta_{-i}))]$$

- Ex-ante: agents know about common prior on types before game starts

$$\mathbb{E}[u_i(s)] = \sum_{\theta_i \in \Theta_i} p(\theta_i)\mathbb{E}[u_i(s,\theta_i)] = \sum_{\theta \in \Theta} p(\theta)\mathbb{E}[u_i(s,\theta)]$$

- Bayes-NE: strategy profile s^* based on ex-ante best response
  - Any finite Bayesian game has a BNE
- Ex-post equilibrium: $s_i^* \in \operatorname{argmax}_{s_i} \mathbb{E}[u_i(s_i, s_{-i}^*, \theta)]\ \forall i, \theta$

Auctions

- Each agent has private type v_i
- Winner chosen by rv $y_i = \max_{j \neq i} v_j$
  - $G_{y_i}(v) = F(v)^{N-1}$
  - $g_{y_i}(v) = (N-1)f(v)F(v)^{N-2}$
- $\mathbb{E}[p(v_i)] = G_{y_i}(v_i)\mathbb{E}[y_i \mid y_i \leq v_i]$
  - $\frac{\mathrm{d}}{\mathrm{d}b}\beta^{-1}(b) = \frac{1}{\beta'(\beta^{-1}(b))}$
  - In symmetric equilibrium, $\beta(v_i) = b_i$
  - Boundary condition: $\beta(0) = 0$, $\beta(x) = 0\ \forall x > v_i$
- Define symmetric and increasing bidding strategy $b_i = \beta(v_i)$
- Optimal bid maximizes $\mathbb{E}[p(v_i)]$
  - Take first derivative
- Second-price auctions
  - Truthful bidding is weak ex-post equilibrium and unique BNE
  - Expected payment is

$$p(v_i) = P(y_i \leq v_i)G_{y_i}(v_i)^{-1}\int_0^{v_i} y g_{y_i}(y)\,\mathrm{d}y = \int_0^{v_i} y g_{y_i}(y)\,\mathrm{d}y$$

- First-price auctions
  - Optimal bid is $b_i = \operatorname{argmax}_{b \geq 0} G_{y_i}(\beta^{-1}(b))(v_i - b)$
  - Take first derivative w.r.t. b eventually results in

$$\beta(v_i) = G_{y_i}(y)^{-1}\int_0^{v_i} y g_{y_i}(y)\,\mathrm{d}y$$

  - Same expected payment as second-price auctions

Perfect Bayesian equilibrium

- Conditions:
  1. Beliefs μ specified
  2. Sequential rationality: strategies s must be optimal given μ
  3. On-the-path consistency: for any on-equilibrium path, μ must be derived from s according to Bayes' rule
  4. Off-the-path consistency: for any off-equilibrium path, μ must be derived from s according to Bayes' rule whenever possible
- Weak PBE: first three conditions
- Strong PBE: all four conditions

Signaling games

- Informed agent move first to signal some information to uninformed agent
- Sending signal is more costly if it contains false information
- Separating: informed agent sends distinct signals for each type
- Pooling: informed agent sends the same signal for all types
- Semi-separating: informed agent sends distinct signals for some types and the same signal for others

Learning in games

- Safety: guarantees at least minmax value
- Rationality: settle on best response to opponent's strategy if opponents are stationary
- No regret: yield payoff no less than any pure strategy

Fictitious play

- Update belief according to

$$\mu_i^t(a_{-i}) = \frac{\eta_i^t(a_{-i})}{\sum_{a'_{-i}} \eta_i^t(a'_{-i})}$$

where $\eta_i^t(a_{-i})$ is the number of times agent i observed a_{-i} at round t.

- Play best response based on empirical distribution: $a_i^{t+1} = \operatorname{argmax}_{a_i} \mu_i^t(a_i, \mu_i^t)$
- Do not need to know opponent's utilities
- Myopic: maximize current utility without considering future ones
- Converges to pure strategies
  - Let $\{a^t\}$ be a sequence of actions generated by FP
  - If $\{a^t\}$ converges to steady state a^*, then a^* is a PSNE
  - If for some t, $a^t = a^*$ where a^* is a strict NE, then $a^\tau = a^*$ for all $\tau > t$
- Proof of strict
  - Suppose $a^t = a^*$
  - We can write μ as

$$\mu_i^{t+1} = (1-\alpha)\mu_i^t + \alpha a_{-i}^t = (1-\alpha)\mu_i^t + \alpha a_{-i}^*$$

  where $\alpha = \frac{1}{\sum_{a'_{-i}} \eta_i^t(a'_{-i}) + 1}$

  - By linearity of expectation, for all a_i,

$$u_i(a_i, \mu_i^t + 1) = (1-\alpha)u_i(a_i, \mu_i^t) + \alpha u_i(a_i, a_{-i}^*)$$

  - Since a_i^* maximizes both terms, the action a_i^* is played
- Converges time-average to mixed strategy NE s^*:

$$\lim_{T \to \infty} \frac{1}{T}\sum_{t=1}^T \mathbb{1}(a_i^t = a_i) = s_i^*(a_i)$$

- Proof
  - Suppose $\{a^t\}$ converges to s^* in time-average sense but s^* is not NE
  - There is some i, a_i with $s_i^*(a_i) > 0$ s.t. $u_i(a'_i, s_{-i}^*) > u_i(a_i, s_{-i}^*)$
  - Choose $\varepsilon < \frac{1}{2}(u_i(a'_i, s_{-i}^*) - u_i(a_i, s_{-i}^*))$
  - Choose T s.t. for all $t \geq T$, $|\mu_i^t(a_{-i}) - s_{-i}^*(a_{-i})| < \varepsilon / \max_{a'} u_i(a')$ for all a_{-i}, which is possible as μ_i^t approaches s_{-i}^* by assumption
  - Then, for any $t \geq T$, we have

$$u_i(a_i, \mu_i^t) \leq u_i(a'_i, \mu_i^t)$$

  - After sufficiently large t, a_i is never played
  - So as $t \to 0$, $\mu_i^t(a_i) \to 0$, which contradicts $s_i^*(a_i) > 0$

Best response dynamics

- Agents start playing arbitrary actions
- In arbitrary order, agents take turns updating their actions to improve their utility
- Repeat until no agent can improve their utility
- If BRD halts, the strategy is a PSNE

Congestion games

- n agents and m resources
- Congestion cost function $l_j(k)$ for cost of resource j when k agents use it
- $n_j(a) = |\{i \mid j \in a_i\}|$
- $c_i(a) = \sum_{j \in a_i} l_j n_j(a)$
- Agents minimize own cost c_i
- NE exists
  - Take potential function

$$\varphi(a) = \sum_{j=1}^m \sum_{k=1}^{n_j(a)} l_j(k) \quad k \in A$$

  - If this is not NE, some agent i switches a_i to b_i, with

$$\Delta c_i(a) = \sum_{j \in b_i \setminus a_i} l_j(n_j(a)+1) + \sum_{j \in a_i \setminus b_i} l_j(n_j(a)) < 0$$

  - Change in potential is

$$\Delta\varphi(a) = \Delta c_i(a)$$

  - Since potential can only take finite number of values, BRD must halt
- φ is an exact potential function if $\Delta\varphi = \Delta c_i$
- φ is an ordinal potential function if $\Delta c_i < 0 \implies \Delta\varphi < 0$ (same sign)
- BRD is guaranteed to half iff the game has an ordinal potential function

No-regret learning

- N experts make predictions $p_i^t \in \{U, D\}$
- One expert is always correct
- **Halving algorithm**: predict by majority vote, observe true outcome, eliminate all wrong experts
  - Converges in $O(\log N)$
  - Makes at most $\log N$ mistakes
- **Iterated halving algorithm**: reset if no expert remain
  - Works when best expert makes k mistakes
  - Makes at most $(k+1)\log N$ mistakes
- **Weighted majority algorithm**: use weight vector W^t, half weight of wrong experts each iteration
  - When algorithm makes mistake, at least half of the experts are downweighted, so $W^{t+1} = \frac{3}{4}W^t$
  - Suppose algorithm makes M mistakes, then $W^T \leq N\left(\frac{3}{4}\right)^M$
  - Since the best expert makes k mistakes, $\left(\frac{1}{2}\right)^k \leq N\left(\frac{3}{4}\right)^M \implies M \leq 2.4(k + \log N)$
  - Makes at most $2.4(k + \log N)$ mistakes
- **Multiplicative weights algorithm**: downweight by $w_i^{t+1} = w_i^t e^{-\varepsilon l_i^t}$ where l is the loss function

- ε is learning rate
- For any sequence of losses and experts k,

$$\frac{1}{T}\mathbb{E}[L_{\text{MW}}^T] \leq \frac{1}{T}L_k^T + \varepsilon + \frac{\ln(N)}{\varepsilon T}$$

- Setting $\varepsilon = \sqrt{\frac{\ln(N)}{T}}$ gives

$$\frac{1}{T}\mathbb{E}[L_{\text{MW}}^T] \leq \frac{1}{T}\min_k L_k^T + 2\sqrt{\frac{\ln(N)}{T}}$$

- Average loss approaches best expert exactly at rate $\frac{1}{\sqrt{T}}$
- Can be used to play games (experts ⇔ actions, losses ⇔ costs)

- Proof of minimax theorem using MW:
  - Assume utilities are scaled to $[0, 1]$
  - Let v_1, v_2 be minmax and maxmin values respectively
  - Suppose $v_1 = v_2 + \varepsilon$ for some $\varepsilon > 0$
  - Suppose A2 uses MW and A1 plays best response
  - For A2,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[u_1(a_1^t, a_2^t)] \leq \frac{1}{T}\min_{a_2}\sum_{t=1}^{T}\mathbb{E}[u_1(a_1^t, a_2)] + 2\sqrt{\frac{\ln(N)}{T}}$$

  - Let $\overline{s}_1$ be a mixed strategy which puts weight $\frac{1}{T}$ on each action a_1^t

$$\frac{1}{T}\min_{a_2}\sum_{t=1}^{T}u_1(a_1^t, a_2) = \min_{a_2}\sum_{t=1}^{T}u_1(a_1^t, a_2) = \min_{a_2}u_1(\overline{s}_1, a_2)$$

  - By definition $\min_{a_2} u_1(\overline{s}_1, a_2) \leq \max_{s_1}\min_{s_2} u_1(\overline{s}_1, a_2) = v_2$

$$\frac{1}{T}\mathbb{E}[u_1(a_1^t, a_2^t)] \leq v_2 + 2\sqrt{\frac{\ln(N)}{T}}$$

  - As A1 best responds to A2 mixed strategy,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[u_1(a_1^t, a_2^t)] &= \frac{1}{T}\sum_{t=1}^{T}\max_{a_1}u_1(a_1, a_2^t)\\ &= \frac{1}{T}\sum_{t=1}^{T}\min_{s_2}\max_{a_1}u_1(a_1, s_2)\\ &\geq v_1\end{aligned}$$

  - So $v_1 \leq v_2 + 2\sqrt{\frac{\ln(N)}{T}}$
  - Then $\varepsilon \leq 2\sqrt{\frac{\ln(N)}{T}}$
  - Taking T large enough leads to contradiction

- **Exp3 algorithm**:

$u \leftarrow [\frac{1}{N}, ...]$
$w_i \leftarrow 1\ \forall i$
for $t \in [1, T]$
  $W^t \leftarrow \sum_{i=1}^{N} w_i^t$
  $p_i^t \leftarrow w_i^t / W^t\ \forall i$
  $q_i^t = (1-\gamma)p_i^t + \gamma u$
  Choose i_t randomly by distribution q^t
  Observe loss $l_{i_t}^t$
  Set other experts losses $l_i^t \leftarrow 0\ \forall i \neq i_t$
  Calculate scaled losses $\hat{l}_i^t \leftarrow l_i^t / q_i^t\ \forall i$
  $w_i^{t+1} \leftarrow w_i^t \exp\left(-\varepsilon \hat{l}_i^t\right)\ \forall i$

External regret

- $a^1, ..., a^T$ has external regret of $\Delta(T)$ if for every agent i and action a_i',

$$\frac{1}{T}\sum_{t=1}^{T}u_i(a^t) \geq \frac{1}{T}\sum_{t=1}^{T}u_i(a_i', a_{-i}) - \Delta(T)$$

- If $\Delta(T) \in o_T(1)$ then the sequence has no external regret
- External regret measures regret to the best fixed action in hindsight
- If $a^t, ..., a^T$ has ε external regret, then distribution π that picks actions uniformly forms an ε-approximate CCE
- Suppose all agents use MW algorithm to choose between k actions
- After T steps, sequence of outcomes has external regret $\Delta(T) = 2\sqrt{\log k / T}$

Swap regret

- $a^1, ..., a^T$ has swap regret of $\Delta(T)$ if for every agent i and every switching function F_i,

$$\frac{1}{T}\sum_{t=1}^{T}u_i(a^t) \geq \frac{1}{T}\sum_{t=1}^{T}u_i(F_i(a_i), a_{-i}) - \Delta(T)$$

- If $\Delta(T) \in o_T(1)$ then the sequence has no swap regret
- Swap regret measures regret where every action could have been swapped to another action
- If $a^t, ..., a^T$ has ε swap regret, then distribution π that picks actions uniformly forms an ε-approximate CE