6) $p \in P_K$ is a probability vector belonging to a k-dimensional vector space. Therefore any vector which is perpendicular to it, ~~can be represe~~ will have its coefficients satisfy,

$\sum a_j p_j = 0$ as $\langle a, p \rangle = 0$ where $a \in \mathbb{R}^K$ is a vector.

For every vector there exists a normal so such an 'a' exists.

Now if $\hat{x}(i, x) = a_i + \frac{I_{\{i=1\}} x_1}{p_1}$

i.e.,

$$\hat{x}(i, x) = \begin{cases} a_i + \frac{x_1}{p_1} & \text{if } i = 1 \\ a_i & \text{otherwise} \end{cases}$$

$$\therefore \mathbb{E}(\hat{x}(i, x)) = \sum p_i \hat{x}(i, x_i)$$

$$= \sum_{i=1}^{K} \left(p_i a_i + \frac{p_i I_{\{i=1\}} x_1}{p_1} \right)$$

$$= \sum_i p_i a_i + \frac{\sum_i p_i I_{\{i=1\}} x_1}{p_1}$$

$$= 0 + \frac{p_1 x_1}{p_1}$$

$$= x_1$$

$\therefore$ Any 'a' perpendicular to $p \in P_K$ will satisfy the equation.

Scanned by CamScanner

5) $R_n^{track}(\pi, \nu) = \mathbb{E}\left[\sum_{t=1} \max x_{t i} - \sum_{t=1}^{T} x_{t, I_t}\right]$

Now for any policy, at time t, the sum over probabilities of all actions is $\underline{1}$

$\therefore \sum_{i \in [K]} P_t(i) = 1$

As all $P_t(i) \geq 0$, we can say there exists ~~one~~ atleast one $i \in [K]$ such that $P_t(i) \leq 1/K$, for any distribution. — (1)

Let the sequence of rewards be $\{x_t\}$ such that,

$$x_{t,i} = \begin{cases} 1 & \text{when } i \text{ is such that } P_t(i) = \min\limits_{j \in K} \{P_t(j)\} \\ 0 & \text{otherwise} \end{cases}$$

$= \mathbb{E}\left[\sum \max_i x_{t,i} - \sum x_{t, I_t}\right]$,

Now, $\max_i x_{t,i} = 1 \ \forall t$

$\therefore$ We have, $T - \mathbb{E} \sum_t x_{t, I_t}$

$= T - \sum_t \mathbb{E}(x_{t, I_t})$

Now, $x_{t,i} = 0 \ \forall i$, except one where $P(i) = \min\limits_{j \in K} P(j)$

$\therefore \mathbb{E}(x_{t, I_t}) = P_t(i^*)$, where $i^* = \operatorname{argmin}\limits_{j \in K} P(j)$

$= T - \sum_t P_t(i^*) \times 1$

$\geq T - \sum_t \frac{1}{K}$ (from (1))

$= T - T/K$

$= T(1 - 1/K)$

$\therefore R_n(\pi, \nu) \geq T(1 - 1/K)$

Scanned by CamScanner

4) Now,

$$R_T(\pi, \nu) = \max_{i\in[K]} \sum_{t=1}^{T} x_{t,i} - \mathbb{E}\left[\sum_{t=1}^{T} x_{t,I_t}\right]$$

In a deterministic policy, the sequence of actions of the player (i.e., I_t) is fixed. So there might exist an environment which generates a sequence of rewards (x_t) such that,

$$x_{t,i} = \begin{cases} 0 & i = I_t \\ 1 & i = [K] - I_t \end{cases}$$

$$\therefore R_T(\pi, \nu) = \max_{i\in[K]} \sum_{t=1}^{T} x_{t,i} - \mathbb{E}\left[\sum_{t=1}^{T} x_{t,I_t}\right]$$

$$= \max_{i\in[K]} \sum_{t=1}^{T} x_{t,i} \qquad [\because x_{t,I_t} = 0 \;\forall t]$$

Now, $\max_{i\in[K]} \sum_{t=1}^{T} x_{t,i} \geq \sum_{t=1}^{T} x_{tj} \quad \forall j \in [K]$

$$\therefore K \max_{i\in[K]} \sum_{t=1}^{T} x_{t,i} \geq \sum_{j\in[K]} \sum_{t=1}^{T} x_{tj} \qquad \text{[Summing over all inequalities for } j\in[K]\text{]}$$

$$= \sum_{t=1}^{T} \sum_{j\in[K]} x_{tj} \qquad \text{[Interchanging summation]}$$

$$= \sum_{t=1}^{T} (K-1)$$

$$\Rightarrow \max_{i\in[K]} \sum_{t=1}^{T} x_{t,i} \geq T\frac{(K-1)}{K}$$

$$= T\left(1 - \frac{1}{K}\right)$$

$\therefore R_T(\pi, \nu) \geq T(1 - 1/K)$ Hence proved.

Scanned by CamScanner