A Proof of Theorem 1

Theorem 1 (Mean aggregation encourages exploration)**.** *Suppose π and $\{\pi_i\}_{1\leq i\leq K}$ are sampled from $P(\pi)$, then the entropy of the ensemble policy $\hat{\pi}$ is no less than the entropy of the single policy in expectation, i.e., $\mathbb{E}_{\pi_1,\pi_2,\ldots,\pi_K}\left[\mathcal{H}(\hat{\pi})\right] \geq \mathbb{E}_{\pi}\left[\mathcal{H}(\pi)\right]$.*

Proof.

$$
\begin{aligned}
H(\hat{\pi}) - \frac{1}{K}\sum_{k=1}^{K}\mathcal{H}(\pi_k) &= -\sum_a \frac{\sum_{k=1}^{K}\pi_k(a|s)}{K}\log\left(\frac{\sum_{k'=1}^{K}\pi_{k'}(a|s)}{K}\right) \\
&\quad + \frac{1}{K}\sum_{k=1}^{K}\sum_a \pi_k(a|s)\log(\pi_k(a|s)) \\
&= -\frac{1}{K}\sum_{k=1}^{K}\sum_a \pi_k(a|s)\left(\log\left(\frac{\sum_{k'=1}^{K}\pi_{k'}(a|s)}{K}\right) - \log(\pi_k(a|s))\right) \\
&= -\frac{1}{K}\sum_{k=1}^{K}\sum_a \pi_k(a|s)\log\left(\frac{\sum_{k'=1}^{K}\pi_{k'}(a|s)}{K\pi_k(a|s)}\right) \\
&\geq -\frac{1}{K}\sum_{k=1}^{K}\log\left(\sum_a \pi_k(a|s)\frac{\sum_{k'=1}^{K}\pi_{k'}(a|s)}{K\pi_k(a|s)}\right) \qquad \text{(Jensen's inequality)} \\
&= -\frac{1}{K}\sum_{k=1}^{K}\log\left(\frac{1}{K}\sum_a\sum_{k'=1}^{K}\pi_{k'}(a|s)\right) \\
&= -\frac{1}{K}\sum_{k=1}^{K}\log(1) = 0\,. \\
\Rightarrow H(\hat{\pi}) &\geq \frac{1}{K}\sum_{k=1}^{K}\mathcal{H}(\pi_k)
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathbb{E}_{\pi_1,\pi_2,\ldots,\pi_K}[\mathcal{H}(\hat{\pi})] &\geq \mathbb{E}_{\pi_1,\pi_2,\ldots,\pi_K}\left[\frac{1}{K}\sum_{k=1}^{K}\mathcal{H}(\pi_k)\right] \\
&\geq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{\pi_1,\pi_2,\ldots,\pi_K}[\mathcal{H}(\pi_k)] \\
&\geq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{\pi}[\mathcal{H}(\pi)] \\
&\geq \mathbb{E}_{\pi}\left[\mathcal{H}(\pi)\right]\,.
\end{aligned}
$$

□