Regret Analysis of EXP3 and EXP3-IX

1 Regret Bounds for EXP3 and EXP3-IX

Let T be the time horizon and N the number of arms. At each round t, the learner selects an arm $A_t \in [N]$ according to a distribution $P_t = (P_{t1}, \ldots, P_{tN})$. The adversary assigns losses $y_{ti} \in [0, 1]$.

1.1 Mean and Standard Deviation of the IPW Estimator

In EXP3, the importance-weighted estimator is

$$\tilde{y}_{ti} := \frac{\mathbf{1}\{A_t = i\} y_{ti}}{P_{ti}}, \qquad y_{ti} \in [0, 1].$$

Mean. Conditioning on the history $\mathcal{F}_{t-1}$,

$$\mathbb{E}[\tilde{y}_{ti} \mid \mathcal{F}_{t-1}] = y_{ti}.$$

Variance and Standard Deviation. The conditional second moment is

$$\mathbb{E}[\tilde{y}_{ti}^2 \mid \mathcal{F}_{t-1}] = \sum_{a=1}^{N} P_{ta} \frac{\mathbf{1}\{a = i\} y_{ti}^2}{P_{ti}^2} = \frac{y_{ti}^2}{P_{ti}}.$$

Hence the conditional variance is

$$\mathrm{Var}(\tilde{y}_{ti} \mid \mathcal{F}_{t-1}) = \frac{y_{ti}^2}{P_{ti}} - y_{ti}^2 = y_{ti}^2 \left(\frac{1}{P_{ti}} - 1 \right),$$

and the conditional standard deviation is

$$\boxed{\sqrt{\mathrm{Var}(\tilde{y}_{ti} \mid \mathcal{F}_{t-1})} = y_{ti} \sqrt{\frac{1}{P_{ti}} - 1}.}$$

Upper Bound. Since $0 \le y_{ti} \le 1$,

$$\sqrt{\mathrm{Var}(\tilde{y}_{ti} \mid \mathcal{F}_{t-1})} \le \sqrt{\frac{1}{P_{ti}} - 1} \le \frac{1}{\sqrt{P_{ti}}}.$$

1.2 EXP3: Pseudo-Regret Bound

Importance-Weighted Estimator. EXP3 uses the estimator

$$\tilde{y}_{ti} := \frac{\mathbf{1}\{A_t = i\} y_{ti}}{P_{ti}}, \qquad \tilde{L}_{Ti} := \sum_{t=1}^{T} \tilde{y}_{ti}.$$

It satisfies the unbiasedness property

$$\mathbb{E}[\tilde{y}_{ti} \mid \mathcal{F}_{t-1}] = y_{ti}.$$

Mixed Estimated Loss. Define

$$\tilde{L}_T := \sum_{t=1}^{T} \sum_{i=1}^{N} P_{ti} \tilde{y}_{ti}.$$

Then

$$\mathbb{E}[\tilde{L}_T] = \mathbb{E}\left[\sum_{t=1}^{T} y_{tA_t}\right].$$

Potential Inequality. Using the sub-Gaussian lower-tail inequality for nonnegative random variables, for each t,

$$\sum_{i=1}^{N} P_{ti} \tilde{y}_{ti} \leq \frac{\eta}{2} \sum_{i=1}^{N} P_{ti} \tilde{y}_{ti}^2 - \frac{1}{\eta} \log\left(\sum_{i=1}^{N} P_{ti} e^{-\eta \tilde{y}_{ti}}\right).$$

Summing over $t = 1, \ldots, T$ and using the telescoping identity induced by the EXP3 weights, we obtain for any arm i,

$$\tilde{L}_T - \tilde{L}_{Ti} \leq \frac{\log N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{j=1}^{N} \tilde{y}_{tj}.$$

Bounding the Second Moment Term. Since exactly one arm is selected at each round,

$$\sum_{j=1}^{N} P_{tj} \tilde{y}_{tj}^2 = \frac{y_{tA_t}^2}{P_{tA_t}} \leq \frac{1}{P_{tA_t}},$$

and thus

$$\mathbb{E}\left[\sum_{j=1}^{N} P_{tj} \tilde{y}_{tj}^2 \mid \mathcal{F}_{t-1}\right] = \sum_{j=1}^{N} P_{tj} \frac{1}{P_{tj}} = N.$$

Therefore,

$$\mathbb{E}\left[\sum_{t=1}^{T} \sum_{j=1}^{N} P_{tj} \tilde{y}_{tj}^2\right] \leq TN.$$

Pseudo-Regret Bound. Taking expectations and using $\mathbb{E}[\tilde{L}_{Ti}] = L_{Ti}$, we obtain

$$\mathbb{E}\left[\sum_{t=1}^{T} y_{tA_t}\right] - \min_i \sum_{t=1}^{T} y_{ti} \leq \frac{\log N}{\eta} + \frac{\eta TN}{2}.$$

Choosing

$$\eta = \sqrt{\frac{2 \log N}{TN}},$$

we conclude

$$\boxed{R_T \leq \sqrt{2TN \log N}.}$$

1.3 EXP3-IX: Deterministic Part of Regret Bound

IX Estimator. EXP3-IX uses

$$\hat{y}_{ti} := \frac{\mathbf{1}\{A_t = i\} y_{ti}}{P_{ti} + \gamma}, \qquad \hat{L}_{Ti} := \sum_{t=1}^{T} \hat{y}_{ti}.$$

Define the mixed estimated loss

$$\hat{L}_T := \sum_{t=1}^{T} \sum_{i=1}^{N} P_{ti} \hat{y}_{ti}.$$

1.4 Loss Quantities and Their Meanings

$\tilde{L}_T$: Algorithm's actual cumulative loss

$$\boxed{\tilde{L}_T := \sum_{t=1}^{T} y_{tA_t}}$$

L_{Ti} : Actual cumulative loss of arm i

$$\boxed{L_{Ti} := \sum_{t=1}^{T} y_{ti}}$$

$\hat{L}_T$: Mixed estimated cumulative loss

$$\boxed{\hat{L}_T := \sum_{t=1}^{T} \sum_{j=1}^{N} P_{tj} \hat{y}_{tj}}$$

$\hat{L}_{Ti}$: Estimated cumulative loss of arm i

$$\boxed{\hat{L}_{Ti} := \sum_{t=1}^{T} \hat{y}_{ti}}$$

Regret Decomposition. For any arm i,

$$\tilde{L}_T - L_{Ti} = (\tilde{L}_T - \hat{L}_T) + (\hat{L}_T - \hat{L}_{Ti}) + (\hat{L}_{Ti} - L_{Ti}).$$

IX Bias Term. Using $\mathbf{1}\{A_t = i\} y_{ti} = (P_{ti} + \gamma)\hat{y}_{ti}$, we obtain

$$\tilde{L}_T - \hat{L}_T = \gamma \sum_{j=1}^{N} \hat{L}_{Tj}.$$

Potential Bound. The EXP3 potential argument yields

$$\hat{L}_T - \hat{L}_{Ti} \leq \frac{\log N}{\eta} + \frac{\eta}{2} \sum_{j=1}^{N} \hat{L}_{Tj}.$$

Combined Bound. Combining the above inequalities,

$$\tilde{L}_T - L_{Ti} \leq \frac{\log N}{\eta} + \left(\gamma + \frac{\eta}{2}\right) \sum_{j=1}^{N} \hat{L}_{Tj} + (\hat{L}_{Ti} - L_{Ti}).$$

Choosing $\gamma = \eta/2$, we obtain

$$\boxed{\tilde{L}_T - L_{Ti} \leq \frac{\log N}{\eta} + \eta \sum_{j=1}^{N} \hat{L}_{Tj} + (\hat{L}_{Ti} - L_{Ti})}.$$

The final term $(\hat{L}_{Ti} - L_{Ti})$ corresponds to the estimation error and requires a separate probabilistic analysis.