Consider a two-state Markov chain with states $\{0, 1\}$ and transition matrix $P$. Without loss of generality, let the Markov chain start in state 0, and in each discrete time slot it makes a transition following $P$. At each time slot $t$, let the true state of the Markov chain be $s(t)$ while the estimated state (needs to be output by an algorithm $\mathcal{A}$ for each $t$) be $\hat{s}(t)$. Then the loss at time slot $t$ is $f(s(t), \hat{s}(t))$ for some 'nice' function $f$. For example $f(x, y) = |x - y|^2$, the squared-loss which in this case is either 0 or 1.

An algorithm $\mathcal{A}$ also has the ability to probe the Markov chain at any time slot $t$ by paying a cost of $c$. Thus, if $\mathcal{A}$ probes at time slot $t$, then $s(t)$ is exactly revealed. Under this model, the objective is to find the optimal probe times $\tau_1, \tau_2, \dots$ so that the following cost is minimized

$$\frac{1}{T}\left(\sum_{t=1}^{T} \mathbb{E}\{f(s(t), \hat{s}(t)) + kc\right)$$

where $k = \max_{\tau_i \leq T} i$ the maximum number of probes made by $\mathcal{A}$.

The answer we expect is that if on a probe at time $t$, the state is $s$ then the time to probe next is $T_s$ (which could be random).

**Potential Algorithm $\mathcal{A}$:** Consider that the current slot is $t$ and the last slot at which sampling was done is $t^-$. Then if there exists a $t'$ such that $t^- < t' \leq t$ for which

$$c + g_s(t', t) \leq g(t', t), \tag{1}$$

where $g_s(t', t)$ is the expected accumulated loss between slots $t'$ and $t$ if sampling is done at time $t'$, while $g_s(t', t)$ is the expected accumulated loss between slots $t'$ and $t$ when sampling is done at time $t^-$, then sample at time $t$.

Claim: This algorithm is 2-approximate.

sub-claim 1: If the algorithm samples at time $t_i$'s, then without loss of generality we can assume that the $\mathsf{OPT}$ samples at least once within each interval $[t_i, t_{i+1}]$. In particular at times $t_i'$.

sub-claim 2: In light of sub-claim 1, the cost of the algorithm is at most

$$c \cdot n_{\mathsf{OPT}} + \text{loss}_{\mathsf{OPT}} + \text{loss}_{\mathcal{A}\setminus\mathsf{OPT}},$$

where $n_{\mathsf{OPT}}$ is the number of times $\mathsf{OPT}$ samples, and $\text{loss}_{\mathcal{A}\setminus\mathsf{OPT}}$ is the loss accumulated by $\mathcal{A}$ but not by $\mathsf{OPT}$ (easiest to see via a picture where at each slot we have a step function indicating the total accumulated loss).

Using the definition of the algorithm (easiest to see pictorially), one can show that $\text{loss}_{\mathcal{A}\setminus\mathsf{OPT}} \leq c n_{\mathsf{OPT}}$.

Algorithm $\mathcal{A}$ is a hindsight algorithm and incurs a loss between $t'$ and $t$ since it can only sample at $t$ even though it would have been beneficial to do so at time $t'$ itself. Given that all the information is stochastic here is an attempt to improve upon $\mathcal{A}$.

Algorithm $\mathcal{A}^c$: At the most recent sampling instant $t^-$, compute the **earliest time** $t > t^-$ for which a $t^- < \tau < t$ exists satisfying (1). Sample at time $\tau$.

**Conjecture:** $\mathcal{A}^c$ **is optimal.**