

Let $q \in [0, 1]^N$ be a fixed vector where $q_i = P(X_i = 1)$. In each time-step t upto a known end point T :

1. The learner chooses an $I_t \in 1, \dots, N$ and $J_t \in 0, 1$.
2. Then $X_t \in 0, 1^N$ is sampled from a product of Bernoulli distributions, $X_{t,i} \sim (q_i)$
3. The learner observes $\tilde{X}_t \in 0, 1^K$, which is defined by $\tilde{x}_{t,i} = \begin{cases} X_{t,i} & \text{if } i \neq I_t \\ J_t & \text{otherwise} \end{cases}$. The learner receives reward $Y_t \sim (r(\tilde{X}_t))$ where $r : 0, 1^K \rightarrow [0, 1]$ is unknown and arbitrary.

