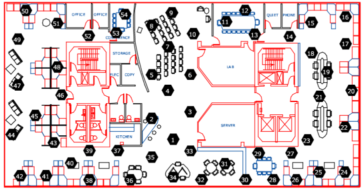
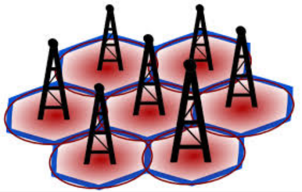
Correlated bandits or: How to minimize mean-squared error online

V. Praneeth Boda[1] and Prashanth L. A.[2]

[1]LinkedIn Corp.
[2]Indian Institute of Technology Madras.

A portion of this work was done while the authors were at University of Maryland, College Park





- Placement of sensors used for measuring temperature in a region.
- Best set of towers which approximate the whole network.

Aim: Find arm with highest information about other arms

Minimum Mean Squared Error Estimation

- Jointly Gaussian arms $X_{\mathcal{M}} = (X_1, \ldots, X_K)$, with zero mean and covariance matrix $\mathbf{\Sigma} \triangleq \mathbb{E}[X_{\mathcal{M}}^T X_{\mathcal{M}}]$.

MMSE

$$\mathcal{E}_i \triangleq \min_g \mathbb{E}\big[\big(X_{\mathcal{M}} - g(X_i)\big)^T \big(X_{\mathcal{M}} - g(X_i)\big)\big]$$
$$= \sum_{j=1}^{K} \mathbb{E}\left[\left(X_j - \mathbb{E}[X_j|X_i]\right)^2\right] = \sum_{j\neq i} \sigma_j^2(1 - \rho_{ij}^2)$$

The optimal

$$g^*(X_i) = \mathbb{E}[X_{\mathcal{M}}|X_i] = [\mathbb{E}[X_1|X_i] \ldots \mathbb{E}[X_K|X_i]]^T,$$
$$\text{with } \mathbb{E}[X_j|X_i] = \frac{\mathbb{E}[X_j X_i]}{\mathbb{E}[X_i^2]} X_i = \frac{\rho_{ij}\sigma_j}{\sigma_i} X_i.$$

Correlated Bandits

Input: set of arm-pairs $\mathcal{S} \triangleq \{(i,j) \mid i,j = 1,\ldots,K, i < j\}$, number of rounds n

based on sample-based MSE-value estimates

For $t = 1, 2, \ldots, n$ **do**

Select a pair $(i_t, j_t) \in \mathcal{S}$

Observe a sample from the bivariate distribution corresponding to the arms i_t, j_t

endfor

necessary for estimating correlation structure

Output an arm $\hat{A}_n$

so that $\mathbb{P}(A_n \neq i^*)$ is minimized.
Here $i^* = \arg\min_{i \in \mathcal{M}} \mathcal{E}_i$.

Based on samples of the Gaussian arms:

MSE of arm i

$$\hat{\mathcal{E}}_i \triangleq \sum_{j \neq i} \hat{\sigma}_j^2 \left(1 - \hat{\rho}_{ij}^2\right).$$

Sample variance

Sample correlation

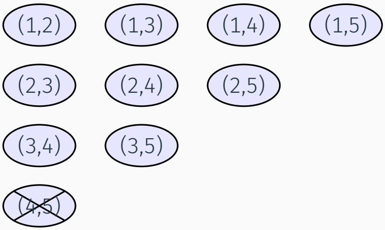
MSE Concentration: Assume $\sigma_i^2 \leq 1, i = 1, \ldots, K$. Then, for any $i = 1, \ldots, K$, and for any $\epsilon \in [0, 2K]$, we have

$$\mathbb{P}\left(\left|\hat{\mathcal{E}}_i - \mathcal{E}_i\right| > \epsilon\right) \leq 14K \exp\left(-\frac{nl^2\epsilon^2}{cK^5}\right),$$

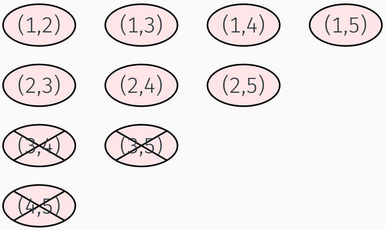
where c is a universal constant, and $0 < l = \min_i \sigma_i^2$.

SR algorithm: Illustration of arm-pair elimination

Maintain active arms and arm-pairs



Active arm-pairs after arms $4, 5$ are eliminated



Active arm-pairs after arms $3, 4, 5$ are eliminated

Successive Rejects: An algorithm to find the best arm

- **Initialization**: $A_1 =$ all arm pairs, $B_1 = \{1, \ldots, K\}$, $n_k = \left\lceil \frac{n - \binom{K}{2}}{C(K)\,(K+1-k)} \right\rceil$, $C(K) \approx K \log K$.
- **Phase 1**: Pull each pair in A_1, n_1 times; Set $B_{k+1} = B_k \backslash$ arm with lowest MSE
- **Phase 2**: Play each arm pair in A_2, $n_2 - n_1$ times; Eliminate ...
- $\vdots$ — $\vdots$
- **Phase $K-1$**: Play the remaining two arm pairs $n_{K-1} - n_{K-2}$ times

- One arm pair played n_1 times, ..., another two played n_2 times
- k arms played n_{k+1} times
- $\sum_{k=1}^{K-1} (k-1) n_k + (K-1) n_{K-1} < n$,
- n_k increases with k
- Adaptive exploration: better than uniform (= play each arm-pair $n / \binom{K}{2}$ times)

Thanks. Questions?