Report on my work with Adobe **Spring 2018**

Report on my work with Adobe

Adobe Advisor: Branislav Kveton *Intern: Subhojyoti Mukherjee*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this communication only with the permission of the Advisor.*

1 Summary Of Discussion

I have been doing literature survey on the area chosen: **Learning latent variable models through online learning**.

2 Problem Definition: Latent variable model

3 Notations

T denotes the time horizon. $\mathcal{A}$ denotes the set of arms with individual arm is denoted by i such that $i = 1, \ldots, K$. The term S_i and F_i denotes the success and the failure respectively for the arm i . The distribution associated with individual arm i is denoted by D_i whereas the reward drawn from that distribution for the t-th time instant is denoted by $x_{i,t}$. r_i denotes the expected mean of the reward distribution D_i and $\hat{r}_i(t)$ denotes the estimated mean for the arm i. All rewards are bounded in $[0, 1]$. n_i denotes the number of times arm i has been pulled.

We define each expert or forecaster as $f_j \in \mathcal{M}_t$, where $\mathcal{M}_t$ is the set of all forecasters at time t. $\mathcal{M}^+{}_t$ is the set of new forecasters introduced at time t. Also, we define $\hat{L}_{f_i,t} = \sum_{s=1}^{t} \ell_{f_i,s}$ as the true cumulative loss suffered by the expert f_i till t-th timestep and $\hat{L}_{f_i,t}$ as the estimated cumulative loss suffered by an expert f_i till t-th timestep such that $\hat{L}_{f_i,t} = \sum_{s=1}^{t} \hat{\ell}_{f_i,s}$. Similarly, we define L_i and $\hat{L}_i$ as the true loss and the estimated loss suffered by arm i. The weight of an expert f_i at time t is defined as $w_{f_i,t}$ and η is defined as a parameter for exploration. Also let $i_{f_j \in \mathcal{M},t}$ be the action suggested by expert f_j at time t.

References

Agrawal, S. and Goyal, N. (2012). Analysis of thompson sampling for the multi-armed bandit problem. In *COLT*, pages 39–1.

Agrawal, S. and Goyal, N. (2013). Further optimal regret bounds for thompson sampling. In *Aistats*, pages 99–107.

Garivier, A. and Moulines, E. (2011). On upper-confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory*, pages 174–188. Springer.

Mellor, J. and Shapiro, J. (2013). Thompson sampling in switching environments with bayesian online change point detection. *CoRR, abs/1302.3721*.