Bocconi

Università Commerciale Luigi Bocconi

# Contextual Thompson Sampling via Generation of Missing Data


## Kelly W. Zhang

Imperial College London


Thursday March 19, 2026
12:00 pm **Room 3-E4-SR03** Via Roentgen 1 Milano


**Abstract**

We introduce a framework for Thompson sampling (TS) contextual bandit algorithms, in which the algorithm's ability to quantify uncertainty and make decisions depends on the quality of a generative model that is learned offline. Instead of viewing uncertainty in the environment as arising from unobservable latent parameters, our algorithm treats uncertainty as stemming from missing, but potentially observable outcomes (including both future and counterfactual outcomes). If these outcomes were all observed, one could simply make decisions using an "oracle" policy fit on the complete dataset. Inspired by this conceptualization, at each decision-time, our algorithm uses a generative model to probabilistically impute missing outcomes, fits a policy using the imputed complete dataset, and uses that policy to select the next action. We formally show that this algorithm is a generative formulation of TS and establish a state-of-the-art regret bound. Notably, our regret bound depends on the generative model only through the quality of its offline prediction loss and applies to any method of fitting the "oracle" policy.



Department of Decision Sciences
Via Röntgen 1 – 20136 Milano

Tel. 02 5836.5632
Fax 02 5836.5630