

Dealing with missing data in Bayesian models for adaptive N-of-1 trials

University of Potsdam

Maliha Raihan Pranti

December 6, 2023

Contents

1 Introduction

In the field of medicine and healthcare, it is known that the same treatment or health intervention may affect individuals differently [?]. Thus, there is a need to look into the cause of difference, may it be behavior, gene or biomarker. N-of-1 trials or Single case design (SCD) is a clinical trial approach that differs from cohort-based trials. In this design, a single subject will serve as one's own control, such that intervention to control comparisons are performed between periods across multiple crossover cycles. Whereas, the traditional randomized controlled trials (RCT) design where the sample population or participants are divided into different groups and then randomly assigned treatments. While N-of-1 trials have historical roots tracing back to the early 20th century [?], their prominence as a patient-centered tool for clinical decision-making has experienced a recent surge in the era of precision medicine [?]. One of the reasons for the N-of-1 trial not being used is the difficulty of generalizing the result on the population level despite being informative on a personal level. This issue is greatly mitigated by researchers later where it is shown that pooling the parallel data of individuals can represent a similar estimate as an RCT [?].

Some situations are more suitable for N-of-1 trials: due to study population, logistics, sample size constraints, or operational costs. An example of this is a rare disease. An update on the SCD methods is adaptive SCD where the dose intensity or treatment is changed using the analysis of interterm data.

A Bayesian approach to adaptive SCD is the multiarm bandit or Thompson sampling where the treatments are considered as arms and, allocates the next treatment to a patient by updating the posterior distribution of the arms [?]. In this design system, if there is missing data, the update of the posterior distribution will suffer, and getting an optimal treatment may take longer than necessary. Even we may have false optimal results due to missing data [?].

Our goal is to find a reliable approach to deal with missing data in Bayesian adaptive N-of-1 trials.

1.1 Related Work

we look for related literature where they discussed the issue of missing data in Thompson sampling or bandit problem as the adaptive N-of-1 trials design con-

1 Introduction

sidered here is based on them. The solution for missing data here ranges from just a simple mean imputation approach to the Doubly robust estimation.

- The paper [?] thoroughly explored the effect of missing data in the bandit algorithms. They incorporate randomized, semi-randomized, and deterministic bandit algorithms. The result showcases that bandit algorithms focused on exploration tend to assign the arm with missing data with the idea of more exploration, while exploitative algorithms are less likely to select the arm with missing data. Next, they experimented with mean imputation to replace missing values in the response-adaptive procedure. The research did all the experiments for a two-armed bandit problem under null ($H_0 : p_0 = p_1$) and under alternative ($H_1 : p_0 < p_1$) settings as well as MAR (Missing at Random). The mean imputation for missing value was based on the current estimated probability of success $p_{k,t} = \frac{\hat{S}_{k,t}}{\hat{S}_{k,t} + \hat{F}_{k,t}}$. Then a sample was drawn from a Bernoulli distribution with $p_{k,t}$. The outcome showed that the impact of missing data on explorative algorithms could be mitigated by mean imputation. However, it was not the same for the exploitive algorithms. A concern was raised about the biases of mean imputation and future research was suggested for a more advanced imputation method.
- Doubly robust Thompson sampling [?] is a novel algorithm for multi-armed contextual bandit. The authors highlighted that we only observe the reward of the chosen arm. So, the estimation process consists of the observed pair of context and reward. Other arms are generally not observed. The estimation using the observed pairs is not efficient. Thus, the Doubly Robust estimator is introduced to fill the unobserved arms reward with pseudo rewards. They have shown that theoretically, it improves the regret bound compared to other methods. Though they did not directly recommend the method to fill up the missingness in the observed reward, we may be able to use the estimator by replacing the $Y_{a_t}(t)$ (the observed reward of arm a at time t), with previous time points observed reward in the formula given below:

$$Y_i^{DR}(t) = \left\{1 - \frac{I(i = a_t)}{\pi_i(t)}\right\} X_i(t)^T \check{\beta}_t + \frac{I(i = a_t)}{\pi_i(t)} Y_{a_t}(t) \quad (1.1)$$

- The authors have considered a scenario where the rewards can be missing in a contextual bandit problem [?]. As an example, a scenario has been described where after giving a treatment, the patient may not come back to give the feedback. In this contextual bandit problem, patient feedback is the reward and the medical record of the patient is context. The proposed method (MLINCUB) adds an online clustering step to fill in the missing

1 Introduction

data in a LINUCB algorithm [?]. With the help of clustering, a reward is drawn from the existing rewards with similar contexts. If the context vector is $x(t)$ and clustered into N clusters, the average reward of each cluster is calculated as follows:

$$\bar{r}_j = \frac{\sum_{\tau=1}^{n_j} r_\tau}{n_j} \quad (1.2)$$

Considering the context of the current iteration is x_t and the reward is missing, the weighted average of m closest clusters reward is assigned as a replacement. The formula to calculate the weighted average is as below:

$$g(x_t) = \frac{\sum_{j=1}^m \frac{\bar{r}_j}{d_j}}{\sum_{j=1}^m \frac{1}{d_j}} \quad (1.3)$$

where d_j represents the distance between x_t and the centroid of j^{th} cluster. The value of m can be customized.

To showcase the efficiency of MLINCUB, they ran LINUCB and MLINCUB on four different datasets with missingness was 10%, 50% and 75%. Except for one dataset, all experiments showed better performance with MINCLUB with a little hyperparameter optimization of N . They added 2D principal component analysis before clustering the context vector as well.

- MM-PGPE(GP) [?] is an approach based on the Monte Carlo implementation of the E step of the EM algorithm [?]. This method is not particularly for Thompson sampling process, but rather generalized for all model-based reinforcement learning. The authors handled the missing data by defining the expectation of the conditional log-likelihood estimate of the target parameters in the M step. The expectation is approximated by averaging over samples drawn from the conditional distribution of missing data with respect to observed data and the parameter obtained in the old M-step. In the E step, a sample is drawn from the conditional distribution using the Metropolis-Hastings algorithm [?]. The experiments done using this method to handle missing data showed better reinforcement learning than the conventional.
- Random marker

Bibliography

- [1] Xijin Chen, Kim May Lee, Sofia S. Villar, and David S. Robertson. Some performance considerations when using multi-armed bandit algorithms in the presence of missing data. *PLoS ONE*, 17, 9 2022. Method: using a simple mean imputation approach.
- [2] Wonyoung Kim, Gi-Soo Kim, and Myunghee Cho Paik. Doubly robust thompson sampling with linear payoffs. While Kim and Paik [2019] uses Lasso estimator with pseudo-rewards aggregated over allarms, we use ridge regression estimator with pseudo-rewards in (1) which are defined separately foreach $i = 1, \dots, N$.
- [3] Djallel Bouneffouf, Sohini Upadhyay, and Yasaman Khazaeni. Contextual bandit with missing rewards. *arXiv preprint arXiv:2007.06368*, 7 2020.
- [4] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.
- [5] Nobuhiko Yamaguchi, Osamu Fukuda, and Hiroshi Okumura. Model-based reinforcement learning with missing data. In *Proceedings - 2020 8th International Symposium on Computing and Networking Workshops, CANDARW 2020*, pages 168–171. Institute of Electrical and Electronics Engineers Inc., 11 2020.
- [6] Greg C. G. Wei and Martin A. Tanner. A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85:699–704, 9 1990.
- [7] Peter Miiller. A generic approach to posterior integration and gibbs sampling. *Tech-nical Report*, pages 91–09, 1991.