Tighter Bounds on Bias Estimation in Doubly Robust Estimators

Sunkai Lyu
Department of Biostatistics, School of Public Health, Peking University
lyusunkai@stu.pku.edu.cn

July 25, 2025

Abstract

Recommender systems have become ubiquitous in personalized service platforms, yet their performance suffers from selection bias—a systemic distortion arising from non-random missing ratings where users preferentially engage with preferred items. While Doubly Robust (DR) estimators have emerged as a dominant solution by concurrently addressing bias and variance, recent studies reveal that conventional bias relaxation techniques adopt excessively coarse approximations, leading to significant overestimation of model bias. This work introduces a novel conservative bias relaxation framework that derives tighter error bounds through theoretical analysis with Lagrange's Identity, and empirically validates lower bias overestimation on an ML100K-based semi-synthetic dataset. The effectiveness of bias correction in practical algorithms is systematically validated on two real-world datasets.

1 Introduction

Recommender systems are now widely deployed to deliver personal recommendations across diverse domains including e-Commerce and social media. In such systems, users rate self-selected items, creating nonrandom missing data due to preference-based selection. This induces distributional shift between observed and complete ratings, introducing selection bias to the models.

To address this issue, three three principled debiasing methodologies have been developed in literature. Error-imputation-based (EIB) [Steck2010] methods attempt to impute missing ratings with a specific model, but heavily rely on the correct model specification. Inverse propensity scoring (IPS) [Saito *et al.*2020, Wang *et al.*2022] methods reweight observed data with inverse propensity. They propose unbiased estimators if the propensity model is correctly specified but often exhibit substantial variance. Doubly robust (DR) [Saito2020, Wang *et al.*2019] methods combine EIB with IPS, guaranteeing unbiasedness when either the imputation model or the propensity score model is correct. The better robustness makes it the preferred approach recently.

However, conventional DR methods inherit the high variance problem from the IPS component, especially when dealing with small propensity scores, which indicates that a slight misspecification of the models may lead to substantial error. [Kang and Schafer2007] To address the limitation, more robust approaches are proposed such as more robust doubly robust (MRDR) [Guo *et al.*2021] and DR-MSE [Dai *et al.*2022] methods. These methods jointly optimize bias-variance trade-offs through penalizing the loss function of the imputation model. Alternatively, Zhou et al. [Zhou *et al.*2023] pioneered a paradigm shift by proposed a generalized propensity learning (GPL) framework to optimize the bias-variance term when learning the propensity model instead of the imputation model. Their work is impressive, though the upper bound of the bias becomes much too loose when they apply the Cauchy-Schwarz Inequality. This work is going to propse a tighter relaxation with Lagrange's Identity. With the tighter upper bound, the bias will be less overestimated.

2 Preliminaries

Let $\mathcal{U} = \{u_1, \ldots, u_m\}$ be the set of users, $\mathcal{I} = \{i_1, \ldots, i_n\}$ be the set of items, and $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ be the set of all pairs of user-items. The rating matrix is denoted as $\mathbf{R} \in \mathbb{R}^{m \times n}$ with $r_{u,i}$, which indicates the rating of the user u on the item i. Let $o_{u,i} \in \{0, 1\}$ be the observation indicator that indicates whether $r_{u,i}$ is observed, and let $\mathbf{x}_{u,i}$ be the associated feature vector. We denote the prediction model as $f_\theta(\cdot)$ parameterized by θ, and the predicted ratings as $\hat{r}_{u,i} = f_\theta(\mathbf{x}_{u,i})$.

The goal is to accurately predict $r_{u,i}$ for all user-item pairs, which can be achieved by minimizing the ideal loss:

$$\begin{aligned}\mathcal{L}_{\text{ideal}}(\theta) &= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} \mathcal{L}(f_\theta(\mathbf{x}_{u,i}), r_{u,i}) \\ &:= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} e_{u,i},\end{aligned} \tag{1}$$

where $\mathcal{L}(\cdot,\cdot)$ is the training loss function, such as cross-entropy loss.

However, we cannot observe the complete rating matrix. Let $\mathcal{O} = \{(u,i)|o_{u,i} = 1\}$ denote the set of user-item pairs with observed ratings. Thus, the naive method minimizes the average loss over the observed samples:

$$\mathcal{L}_N(\theta) = \frac{1}{|\mathcal{O}|}\sum_{(u,i)\in\mathcal{O}} e_{u,i}. \tag{2}$$

Due to selection bias, we have $\mathbb{E}\{\mathcal{L}_N(\theta)\} \neq \mathcal{L}_{\text{ideal}}(\theta)$. Several methods have been proposed to unbiasedly estimate the ideal loss, including EIB, IPS, DR, and their variants. Since EIB and IPS can be regarded as special cases of DR, we focus on DR methods here. The loss function of the vanilla DR method is formulated as:

$$\mathcal{L}_{\text{DR}}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}}\right], \tag{3}$$

where $\hat{p}_{u,i}$ is the estimated propensity score for the true exposure probability $p_{u,i} = \Pr(o_{u,i} = 1|\mathbf{x}_{u,i})$, and $\hat{e}_{u,i}$ is the error from the imputation model $m(\mathbf{x}_{u,i};\phi)$, i.e., $\hat{e}_{u,i} = \mathcal{L}(m(\mathbf{x}_{u,i};\phi), \hat{r}_{u,i})$.

The DR estimator still has high variance due to the form of inverse propensity. Many researchers are dedicated to reducing the variance through various regularization methods. Other researchers notice that the bias of DR estimator also needs to be controlled. Whatever the detailed methods are, they are related to the two terms: The squared bias of the DR estimator:

$$\text{Bias}^2[\mathcal{L}_{\text{DR}}(\theta)] = \frac{1}{|\mathcal{D}|^2}\left[\sum_{(u,i)\in\mathcal{D}}\frac{(\hat{p}_{u,i}-p_{u,i})}{\hat{p}_{u,i}}(e_{u,i}-\hat{e}_{u,i})\right]^2. \tag{4}$$

And the variance of the DR estimator is:

$$\text{Var}[\mathcal{L}_{\text{DR}}(\theta)] = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\frac{p_{u,i}-p_{u,i}^2}{\hat{p}_{u,i}^2}(e_{u,i}-\hat{e}_{u,i})^2 \tag{5}$$

In this paper we will focus on the bias. The conventional relaxation for the squared bias is based on Cauchy-Schwarz Inequality:

$$\begin{aligned}&\text{Bias}^2[\mathcal{L}_{\text{DR}}(\theta)] \\ &= \frac{1}{|\mathcal{D}|^2}\left[\sum_{(u,i)\in\mathcal{D}}\frac{(\hat{p}_{u,i}-p_{u,i})}{\hat{p}_{u,i}}(e_{u,i}-\hat{e}_{u,i})\right]^2 \\ &\leq \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}1^2\cdot\sum_{(u,i)\in\mathcal{D}}\frac{(p_{u,i}-\hat{p}_{u,i})^2}{\hat{p}_{u,i}^2}(e_{u,i}-\hat{e}_{u,i})^2 \\ &= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{p_{u,i}^2-2p_{u,i}\hat{p}_{u,i}+\hat{p}_{u,i}^2}{\hat{p}_{u,i}^2}(e_{u,i}-\hat{e}_{u,i})^2\end{aligned} \tag{6}$$

Denote $g_{u,i} = \frac{(\hat{p}_{u,i}-p_{u,i})}{\hat{p}_{u,i}}(e_{u,i}-\hat{e}_{u,i})$, the equation is satisfied if and only if $g_{u,i}$ is all the same, which is nearly impossible in real data, which leads to great gap between the upper bound and the real bias.

3 Methods

3.1 Related works

As we have briefly reviewed above, many researchers attempted to control the variance and the bias of DR estimator by adding regularization terms while optimizing the imputation model. In the work of more robust doubly robust methods, they proposed to add the variance term. The bias term is also included in methods like DR-bias and DR-MSE [Dai *et al.*2022].

Zhou et al.[Zhou *et al.*2023] reviewed the existing methods and found that they basically focused on the imputation model and did not pay enough attention to the propensity model which is an equally important factor which influences the bias and variance of DR-estimator. They proposed the generalized propensity learning framework, which suggested the joint learning of the propensity model and imputation and prediction model.

However, when they tried to estimate the bias of the DR estimator, they all simply take the conventional form. For instance, in GPL framework, they relax the Bias term as:

$$\begin{aligned}&\text{Bias}^2[\mathcal{L}_{\text{DR}}(\theta)] \\ &= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{p_{u,i}^2-2p_{u,i}\hat{p}_{u,i}+\hat{p}_{u,i}^2}{\hat{p}_{u,i}^2}(e_{u,i}-\hat{e}_{u,i})^2 \\ &\leq \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{p_{u,i}-2p_{u,i}\hat{p}_{u,i}+\hat{p}_{u,i}^2}{\hat{p}_{u,i}^2}(e_{u,i}-\hat{e}_{u,i})^2 \\ &= L_{bias}\end{aligned} \tag{7}$$

In the following parts we are going to prove this approximation brings about large overestimation, which may cause the optimizer to pay too much attention to the bias term.

3.2 Main theory

Our work is aimed to reduce the overestimation of the bias. Recall form of the squared bias:

$$\begin{aligned}\text{Bias}^2[\mathcal{L}_{\text{DR}}(\theta)] &= \frac{1}{|\mathcal{D}|^2}\left[\sum_{(u,i)\in\mathcal{D}} \frac{(\hat{p}_{u,i}-p_{u,i})}{\hat{p}_{u,i}}(e_{u,i}-\hat{e}_{u,i})\right]^2 \\ &= \frac{1}{|\mathcal{D}|^2}(\sum_{(u,i)\in\mathcal{D}} g_{u,i})^2.\end{aligned} \tag{8}$$

To analyze the gap of Cauchy-Schawarz Inequality, we need Lagrange's Identity:

Lemma 1 (Lagrange's Identity)**.**

$$\begin{aligned}\left(\sum_{k=1}^{n} a_k b_k\right)^2 &= \left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right) \\ &- \sum_{1\leq i<j\leq n}(a_i b_j - a_j b_i)^2, \forall \mathbf{a},\mathbf{b}\in\mathbb{R}^n,\end{aligned} \tag{9}$$

The vector form is:

$$\|\mathbf{a}\|^2\|\mathbf{b}\|^2 = (\mathbf{a}\cdot\mathbf{b})^2 + \|\mathbf{a}\times\mathbf{b}\|^2 \tag{10}$$

Proof. There are many proofs for the identity, we provide one of them:

$$\begin{aligned}\|\mathbf{a}\|^2\|\mathbf{b}\|^2 &= \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{j=1}^{n} b_j^2\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^2 b_j^2 \\ &= \sum_{i=1}^{n} a_i^2 b_i^2 + \sum_{i\neq j} a_i^2 b_j^2 \\ (\mathbf{a}\cdot\mathbf{b})^2 &= \left(\sum_{i=1}^{n} a_i b_i\right)^2 \\ &= \sum_{i=1}^{n} a_i^2 b_i^2 + 2\sum_{i<j} a_i a_j b_i b_j \\ \|\mathbf{a}\times\mathbf{b}\|^2 &= \sum_{i<j}(a_i b_j - a_j b_i)^2 \\ &= \sum_{i<j}(a_i^2 b_j^2 + a_j^2 b_i^2 - 2a_i a_j b_i b_j)\end{aligned}$$

Combining these:

$$\begin{aligned}(\mathbf{a}\cdot\mathbf{b})^2 + \|\mathbf{a}\times\mathbf{b}\|^2 &= \sum_{i=1}^{n} a_i^2 b_i^2 + \sum_{i<j}(a_i^2 b_j^2 + a_j^2 b_i^2) \\ &= \sum_{i=1}^{n} a_i^2 b_i^2 + \sum_{i\neq j} a_i^2 b_j^2 = \|\mathbf{a}\|^2\|\mathbf{b}\|^2\end{aligned}$$

□

Using Lemma 1,

$$\begin{aligned}\text{Bias}^2[\mathcal{L}_{\text{DR}}(\theta)] &= \frac{1}{|\mathcal{D}|^2}(\sum_{(u,i)\in\mathcal{D}} g_{u,i})^2 \\ &= \frac{1}{|\mathcal{D}|^2}(\sum_{(u,i)\in\mathcal{D}} 1^2 \sum_{(u,i)\in\mathcal{D}} g_{u,i}^2 \\ &\quad - \frac{1}{2}\sum_{(u,i)\neq(u',i')}(g_{u,i}-g_{u',i'})^2) \\ &= \underbrace{\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} g_{u,i}^2}_{\text{Conventional Relaxation Term}} \\ &\quad - \frac{1}{2|\mathcal{D}|^2}\sum_{(u,i)\neq(u',i')}(g_{u,i}-g_{u',i'})^2.\end{aligned} \tag{11}$$

The first term is exactly the conventional relaxation, and the second term is the gap we want to discuss.

To deal with the squares of the difference, we need two simple lemmas, which allows us to estimate both the upper bound and the lower bound of the gap:

Lemma 2 (Square Inequality 1)**.** $a \leq b \leq c \in \mathbb{R}$, $(a-b)^2 + (b-c)^2 \leq (a-c)^2$

Proof. The inequality is equivalent to $(a-b)(b-c) \leq 0$, which can be obtained by the condition. □

This inequality helps to determine the upper bound of the gap.

Theorem 1 (Upper bound of the relaxation gap)**.**

$$\sum_{(u,i)\neq(u',i')}(g_{u,i}-g_{u',i'})^2 \leq \frac{|D|^2}{2}(\max_{u,i} g_{u,i} - \min_{u,i} g_{u,i})^2$$

Proof. : Here we rewrite the variables as $g_{u|I|+i} = g_{u,i}$. Now the variables become $g_1, \cdots, g_N$, where $N = |D|$. Without loss of generation, we may assume that $m = g_1 \leq g_2 \leq \cdots \leq g_N = M$. We are going to prove that

$$\sum_{k\neq j}(g_k - g_j)^2 \leq \frac{N^2}{2}(M-m)^2.$$

We adopt the inductive method to prove this statement. If $N=1$, the inequality is trivial. If $N=2$, $(g_1-g_2)^2+(g_2-g_1)^2 \le 2(g_2-g_1)^2$ because they are equal in fact.

Suppose that the inequality holds for $N \le t$, when $N = t+1$,

$$\begin{aligned}\sum_{k\neq j}(g_k-g_j)^2 &= 2(g_1-g_{t+1})^2 \\ &\quad + 2\sum_{k=2}^{t-1}[(g_1-g_k)^2+(g_k-g_{t+1})^2] \\ &\quad + \sum_{2\le k\neq j\le t}(g_k-g_j)^2 \\ &\le 2(g_1-g_{t+1})^2+2(t-1)(g_1-g_{t+1})^2 \\ &\quad + \sum_{2\le k\neq j\le t}(g_k-g_j)^2 \\ &= 2t(g_1-g_{t+1})^2+\sum_{2\le k\neq j\le t}(g_k-g_j)^2\end{aligned}$$

The last line holds because of Lemma 2. Since $m = g_1 \le g_2 \le \cdots \le g_t \le g_{t+1} = M$, according to our inductive assumption,

$$\begin{aligned}&\sum_{1\le k\neq j\le t+1}(g_k-g_j)^2 \\ &\le 2t(g_1-g_{t+1})^2+\sum_{2\le k\neq j\le t}(g_k-g_j)^2 \\ &\le 2t(g_1-g_{t+1})^2+\frac{(t-1)^2}{2}(g_2-g_t)^2 \\ &\le 2t(g_1-g_{t+1})^2+\frac{(t-1)^2}{2}(g_1-g_{t+1})^2 \\ &= \frac{(t+1)^2}{2}(g_1-g_{t+1})^2.\end{aligned}$$

That completes the inductive proof. □

It might be questioned that the upper bound will never be achieved but unfortunately if $|D|$ is even and exact half of $g_{u,i}$s equal the minimum while the other half of them equal the maximum, the equation always holds, which suggests the relaxation here is surprisingly unreliable.

In the next part, we will analyze the lower bound of the gap and propose a modification. We have another lemma:

Lemma 3 (Square Inequality 2)**.** $(a-b)^2+(b-c)^2 \ge \frac{1}{2}(a-c)^2, \forall a,b,c \in \mathbb{R}$

Proof. The inequality is equivalent to $(a+c-2b)^2 \ge 0$, which is trivial. □

With Lemma 3, we may derive the lower bound as:

Theorem 2 (Lower bound of the relaxation gap)**.**

$$\begin{aligned}&\sum_{(u,i)\neq(u',i')}(g_{u,i}-g_{u',i'})^2 \\ &\ge \left(\frac{|\mathcal{D}|+|\mathcal{O}|}{2|\mathcal{O}|}\right)\sum_{(u,i)\neq(u',i')\in\mathcal{O}}(g_{u,i}-g_{u',i'})^2\end{aligned}$$

Proof.

$$\begin{aligned}&\sum_{(u,i)\neq(u',i')}(g_{u,i}-g_{u',i'})^2 \ge \sum_{(u,i)\neq(u',i')\in\mathcal{O}}(g_{u,i}-g_{u',i'})^2 \\ &+\frac{1}{|\mathcal{O}|}\sum_{\substack{(u'',i'')\notin\mathcal{O}\\(u,i)\neq(u',i')\in\mathcal{O}}}[(g_{u,i}-g_{u'',i''})^2+(g_{u'',i''}-g_{u',i'})^2] \\ &\ge \sum_{(u,i)\neq(u',i')\in\mathcal{O}}(g_{u,i}-g_{u',i'})^2+\frac{|\mathcal{O}^c|}{2|\mathcal{O}|}\sum_{(u,i)\neq(u',i')\in\mathcal{O}}(g_{u,i}-g_{u',i'})^2 \\ &= \left(\frac{|\mathcal{D}|+|\mathcal{O}|}{2|\mathcal{O}|}\right)\sum_{(u,i)\neq(u',i')\in\mathcal{O}}(g_{u,i}-g_{u',i'})^2\end{aligned}$$

□

Thus we have

$$\begin{aligned}&\text{Bias}^2[\mathcal{L}_{\text{DR}}(\theta)] \\ &\le \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}g_{u,i}^2-\frac{|\mathcal{D}|+|\mathcal{O}|}{2|\mathcal{D}|^2|\mathcal{O}|}\sum_{(u,i)\neq(u',i')\in\mathcal{O}}(g_{u,i}-g_{u',i'})^2.\end{aligned} \tag{12}$$

We call the latter term **Bias correction**. This inequality provides a correction to consturct a tighter bound of bias.

Here we must explain why we should use the $g_{u,i}$s where $(u,i) \in \mathcal{O}$ instead of all $g_{u,i}$s. Recall that $g_{u,i} = \frac{(\hat{p}_{u,i}-p_{u,i})}{\hat{p}_{u,i}}(\hat{e}_{u,i}-e_{u,i})$. $g_{u,i}$ is related to $p_{u,i}$ and $e_{u,i}$. But $e_{u,i}$ is definitely unknown for $o_{u,i}=0$, which makes us focus on the terms where errors can be observed in practice.

Now that we only consider (u,i) where $o_{u,i}=1$, it is improper to treat $o_{u,i}$ as an estimator for $p_{u,i}$. Since $o_{u,i}$ is generated from Bernoulli sampling with probability $p_{u,i}$, $p_{u,i}$ can be estimated by $\bar{o}_{u,i}$, the mean of K-nearest neighbors based on the features of (u,i).

This work makes contributions to many works which require an estimate for the bias of the DR estimator including GPL, DR-bias and DR-MSE. With tighter bound for the bias, these methods are able to balance the empirical loss and the regularization.

4 Experiments

4.1 Semi-synthetic Experiments

4.1.1 Setups

We use ML100K to construct a semi-synthetic dataset to explore the true effect of the tighter bound. To be detailed, we first process the data as follows:

(1) Use matrix factorization to complete the rating matrix. Then we sort all ratings in ascending order and truncated based on the empirical rating distribution from observed data $[p_1, p_2, p_3, p_4, p_5] = [0.0611, 0.1137, 0.2715, 0.3417, 0.2120]$. Set the lowest p_1 fraction to $R_{u,i} = 1$, then the next p_2 fraction to $R_{u,i} = 2$ and so on.

(2) For $R_{u,i} \in \{1, 2, 3, 4, 5\}$, set the CTR $p_{u,i} \in (0, 1)$ with $p_{u,i} = p^{\alpha} max(1, 5 - R_{u,i})$. , where p is set to 1 and α is initially set to 0.5 in our experiments.

(3) Transform the predicted ratings $R_{u,i}$ into true CVR $r_{u,i}$ by correspondingly replacing the rating 1, 2, 3, 4, 5 with the conversion rate 0.1, 0.3, 0.5, 0.7, 0.9, and sample the binary click indicator and conversion label with the Bernoulli sampling $o_{u,i} \sim Bern(p_{u,i})$, $r_{u,i}^{obs} \sim Bern(r_{u,i})$.

(4) Generate the CVR prediction $\hat{r}_{u,i}$ with the following methods:

1. **Method ONE:**

$$\hat{r}_{u,i} = \begin{cases} 0.9 & \text{for randomly selected } \{(u,i)|r_{u,i} = 0.1\} \\ r_{u,i} & \text{otherwise} \end{cases}$$

2. **Method THREE:**

$$\hat{r}_{u,i} = \begin{cases} 0.9 & \text{for randomly selected } \{(u,i)|r_{u,i} = 0.3\} \\ r_{u,i} & \text{otherwise} \end{cases}$$

3. **Method FIVE:**

$$\hat{r}_{u,i} = \begin{cases} 0.9 & \text{for randomly selected } \{(u,i)|r_{u,i} = 0.5\} \\ r_{u,i} & \text{otherwise} \end{cases}$$

4. **Method ROTATE:**

$$\hat{r}_{u,i} = \begin{cases} r_{u,i} - 0.2 & \text{if } r_{u,i} \geq 0.3 \\ 0.9 & \text{if } r_{u,i} = 0.1 \end{cases}$$

5. **Method SKEW:**

$$\hat{r}_{u,i} \sim \mathcal{N}\left(r_{u,i}, \sigma^2\right)$$

where $\sigma = \frac{1 - r_{u,i}}{2}$ and clipped to $[0.1, 0.9]$.

6. **Method CRS:**

$$\hat{r}_{u,i} = \begin{cases} 0.2 & \text{if } r_{u,i} \leq 0.6 \\ 0.6 & \text{otherwise} \end{cases}$$

(5) Other settings:

Set the prediction of the propensity $\hat{p}_{u,i}$ to derive from $\frac{1}{\hat{p}_{u,i}} = \frac{\beta}{p_{u,i}} + \frac{1-\beta}{p_a}$, where $p_a = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} o_{u,i}$. Define the imputed error $\hat{e}_{u,i} = \text{CE}\left(\sum_{(u,i) \in O} r_{u,i}^{obs} w_{u,i}, \hat{r}_{u,i}\right)$, where $w_{u,i} = \frac{1/\hat{p}_{u,i}}{\sum_{(u,i) \in O} 1/\hat{p}_{u,i}}$. CE denotes the cross-entropy loss.

4.1.2 Metrics

In the semi-synthetic experiments, we can derive the accurate **Bias**^{2}. To validate the effect of our new bound, we define two important metrics:

$$\textbf{Relative Improvement} = \frac{\textbf{Bias correction}}{L_{bias} - \textbf{Bias}^2},$$

$$\textbf{To Bias Improvement} = \frac{\textbf{Bias correction}}{\textbf{Bias}^2}.$$

The first metric indicates how much the correction term helps correct the overestimation, while the second metric reflects the correction in real value.

4.1.3 Results

We perform the experiments when $\alpha = 0.3, 0.5, 0.7$. For each α we repeat the experiment 20 times and take the average, the result is shown as follows:

Table 1: Relative Improvement Across Different α Values (Averaged over 20 Runs)

Method	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$
ONE	0.012 759	0.044 000	0.149 097
THREE	0.015 808	0.069 330	0.262 142
FIVE	0.021 437	0.105 818	0.385 252
ROTATE	0.005 120	0.026 176	0.097 384
SKEW	0.018 501	0.087 803	0.321 933
CRS	0.006 162	0.041 619	0.194 400

From the two tables we find that the correction works. And the more data is observed, as α increases, the more powerful the correction functions, and vice versa.

Table 2: To Bias Improvement Across Different α Values (Averaged over 20 Runs)

Method	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$
ONE	0.206 766	0.598 607	2.404 227
THREE	0.233 665	0.805 354	3.963 850
FIVE	0.277 257	1.073 554	6.967 757
ROTATE	0.079 836	0.396 874	1.888 813
SKEW	0.444 988	1.462 412	8.638 609
CRS	0.202 168	1.348 659	23.913 068

4.2 Real-world Experiments

4.2.1 Setups

In this section, we test the performance of bias correction in real-world datasets Coat and Yahoo!R3. These two datasets are widely used in debiased recommender systems because they contain both biased and unbiased data. The Coat dataset comprises evaluation data from 290 users on 300 products, including 6,960 biased ratings in the training set and 4,640 unbiased ratings in the test set. The Yahoo! R3 dataset represents a larger-scale collection, containing 311,704 potentially biased ratings from 15,400 users on 1,000 items in its training portion, along with 54,000 unbiased test ratings sampled from the first 5,400 users. Both datasets originally used a 5-point rating scale, which was subsequently binarized for modeling purposes - ratings below 3 were converted to 0, while others were mapped to 1. This transformation preserves the essential preference relationships while simplifying the prediction task. The datasets' contrasting scales, density patterns, and bias characteristics provide a robust testbed for evaluating model performance across different scenarios.

We use DR-JL, which represents the joint learning of the imputation model and the prediction model, and DR-JL-GPL as baselines. DR-JL-GPL combines GPL framework with DR-JL, which utilized the estimate of L_{bias} in the learning of the propensity model. Our work helps correct the overestimation of L_{bias}, which is denoted as DR-JL-GPL with tighter bounds (DR-JL-GPL-TB).

We implement the three mentioned methods with Pytorch, employing Adam. The learning rate is chosen from $[0.0001, 0.1]$, and decay rate is chosen from $[0.0001, 0.01]$. The regularization coefficient λ of DR-GPL and DR-GPL-TB ranges from $[0.01, 0.1]$. The balancing parameter between bias penalty and variance penalty is set among $\{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ to adapt for the bias penalty after correction.

4.2.2 Metrics

To evaluate prediction performance, we adopt three metrics: AUC, $NDCG@K$ (denoted as $N@K$), and $Recall@K$ (denoted as $R@K$). $N@K$ and $R@K$ are particularly popular in recommender systems as they assess the quality of ranking, which is central to recommendation tasks. The expressions are given as follows:

$$DCG_u@K = \sum_{i \in D_u^{test}} \frac{\mathbb{I}(\hat{z}_{u,i} \leq K)}{\log(\hat{z}_{u,i} + 1)},$$

$$NDCG@K = \frac{1}{|U|} \sum_{u \in U} \frac{DCG_u@K}{IDCG_u@K},$$

$$Recall_u@K = \frac{\sum_{i \in D_{test}^u} \mathbb{I}(\hat{z}_{u,i} \leq K)}{\min(K, |D_u^{test}|)},$$

$$Recall@K = \frac{1}{|U|} \sum_{u \in U} Recall_u@K.$$

where IDCG represents the best possible DCG, D_{test}^u denotes the cardinality of all ratings of the user u in test data, and $\hat{z}_{u,i}$ represents the ranking of item i in the recommended list for user u. We set $K = 5$.

Table 3: Performance on Yahoo! R3 Dataset

Method	AUC	N@5	R@5
DR-JL	0.6853	0.6613	0.4240
DR-JL-GPL	0.6988	0.6711	0.4322
DR-JL-GPL-TB	0.6990	0.6712	0.4315

Table 4: Performance on Coat Dataset

Method	AUC	N@5	R@5
DR-JL	0.7160	0.6752	0.4718
DR-JL-GPL	0.7380	0.6735	0.4643
DR-JL-GPL-TB	0.7403	0.6713	0.4655

The experimental results indicate that DR-JL-GPL-TB demonstrates comparable performance to DR-JL-GPL without showing significant improvement. Although a marginal increase in AUC is observed, other evaluation metrics exhibit inconsistent trends. This is reasonable because in these datasets, the observed data is very sparse, making the bias correction extremely weak. Anyhow, it is merely a modification of DR-GPL. One potential direction to fully exploit its capabilities is to adaptively treat items with minimal error terms as observable entries, thereby enhancing the correction intensity.

5 Conclusions

We have proposed a novel bias correction method that demonstrably improves estimation accuracy across multiple experimental scenarios. Our results show consistent performance gains under varying α parameters, with particularly significant improvements observed in high-α regimes. The proposed correction mechanism successfully addresses the inherent bias in the original estimator while maintaining computational efficiency.

This approach exhibits three key advantages: (1) mathematical tractability, (2) minimal computational overhead, and (3) compatibility with existing evaluation frameworks. Crucially, the methodology is algorithm-agnostic by design, suggesting immediate applicability to a broad spectrum of recommendation algorithms beyond those tested here - including neural collaborative filtering, graph-based methods, and hybrid systems.

Future work may investigate the correction's effectiveness in cold-start scenarios and its integration with deep learning architectures. The consistent performance patterns observed suggest this bias correction could become a standard component in recommendation system evaluation pipelines.

References

[Dai *et al.*2022] Quanyu Dai, Haoxuan Li, Peng Wu, Zhenhua Dong, Xiao-Hua Zhou, Rui Zhang, Rui zhang, and Jie Sun. A generalized doubly robust learning framework for debiasing post-click conversion rate prediction, 2022.

[Guo *et al.*2021] Siyuan Guo, Lixin Zou, Yiding Liu, Wenwen Ye, Suqi Cheng, Shuaiqiang Wang, Hechang Chen, Dawei Yin, and Yi Chang. Enhanced doubly robust learning for debiasing post-click conversion rate estimation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 275–284. ACM, July 2021.

[Kang and Schafer2007] Joseph D. Y. Kang and Joseph L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4), November 2007.

[Saito *et al.*2020] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, page 501–509, New York, NY, USA, 2020. Association for Computing Machinery.

[Saito2020] Yuta Saito. Doubly robust estimator for ranking metrics with post-click conversions. In *Proceedings of the 14th ACM Conference on Recommender Systems*, RecSys '20, page 92–100, New York, NY, USA, 2020. Association for Computing Machinery.

[Steck2010] Harald Steck. Training and testing of recommender systems on data missing not at random. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, page 713–722, New York, NY, USA, 2010. Association for Computing Machinery.

[Wang *et al.*2019] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. Doubly robust joint learning for recommendation on data missing not at random. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6638–6647. PMLR, 09–15 Jun 2019.

[Wang *et al.*2022] Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. Escm2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 363–372. ACM, July 2022.

[Zhou *et al.*2023] Yuqing Zhou, Tianshu Feng, Mingrui Liu, and Ziwei Zhu. A generalized propensity learning framework for unbiased post-click conversion rate estimation. In *CIKM*, pages 3554–3563, 2023.