Off-policy evaluation for learning-to-rank via interpolating the item-position model and the position-based model

ALEXANDER BUCHHOLZ, Amazon Music ML, Germany
BEN LONDON, Amazon Music ML, US
GIUSEPPE DI BENEDETTO, Amazon Music ML, Germany
THORSTEN JOACHIMS, Cornell University, US

A critical need for industrial recommender systems is the ability to evaluate recommendation policies offline, before deploying them to production. Unfortunately, widely used off-policy evaluation methods either make strong assumptions about how users behave that can lead to excessive bias, or they make fewer assumptions and suffer from large variance. We tackle this problem by developing a new estimator that mitigates the problems of the two most popular off-policy estimators for rankings, namely the position-based model and the item-position model. In particular, the new estimator, called INTERPOL, addresses the bias of a potentially misspecified position-based model, while providing an adaptable bias-variance trade-off compared to the item-position model. We provide theoretical arguments as well as empirical results that highlight the performance of our novel estimation approach.

Additional Key Words and Phrases: Off-policy evaluation, Learning-to-rank, Position bias, Position-based model, Item-position model

ACM Reference Format:
Alexander Buchholz, Ben London, Giuseppe Di Benedetto, and Thorsten Joachims. 2022. Off-policy evaluation for learning-to-rank via interpolating the item-position model and the position-based model. In *Proceedings of CONSEQUENCES - RecSys 2022*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Online media streaming platforms rely on highly personalized content recommendation that allows users to navigate large content pools [3, 12]. As the underlying ranking policies constantly evolve, recommendation providers need to experiment offline with new approaches for ranking content before actually deploying and exposing them to the users [5, 6]. This serves the purpose of deploying only policies that have a large chance of improving the user experience. Deployed ranking policies provide a plethora of interaction logs that can be repurposed to learn and evaluate potentially better policies offline. These logs come in the form of implicit feedback, i.e., records of past interaction behavior, linked to information about the user, the context and the items to recommend. Off-policy evaluation of new policies on historic data requires adequate strategies to deal with biases coming from (i) the nature of user interaction and (ii) the logging policy. A prominent example of these biases is position bias [7] (content that is not ranked in the most visible positions is less likely to be seen). We focus on two popular classes of estimators that take different approaches to correcting for presentation bias. The first class, in the case of full visibility of all items, does not rely on explicit randomization, but models the randomness in user behavior. The most common model is the position-based model (PBM), which assumes that observed clicks on content factorize into relevance (depending on the item only) and the visibility of the content (depending on the position only). The PBM has been successfully used in practice and various methods

Manuscript submitted to ACM

have been developed for estimating position bias curves [1, 2, 9, 14, 18]. In realistic scenarios, the position bias curve (i.e., examination probabilities) must be estimated and so will almost always be approximate, which can lead to a bias in the evaluation (even if true user behavior factors as assumed by the PBM). The other class of estimators requires explicit randomization during data collection. The most popular estimator in this class is the item-position model (IPM) [11]. Unlike the PBM, the IPM allows clicks to be a function of interactions between the recommended content and its display position and does not require the estimation of a position bias curve. Since it uses explicit randomization, its bias is typically lower but at the expense of increased estimator variance.

To obtain a better balance of bias and variance under realistic conditions, we propose a new estimator that interpolates between the PBM and the IPM in the full visibility setting. We show that this estimator is always unbiased for a correctly specified PBM, but can have better variance than both the PBM and the IPM. More importantly, for a misspecified PBM, our new estimator is based on the idea that the position-based model might provide a good approximation to local behavior, i.e., small differences in ranking position are properly modeled, but large jumps from the top to the bottom of a list lead to unreliable examination probabilities. The IPM calibrates the PBM by computing probabilities that an item is within the range (i.e., window size) of another one, if the PBM is correctly specified. The window size serves as a tuning parameter that allows to trade off potentially high variance IPM estimation with a potentially more biased PBM. We show empirically that this leads to reduced error and hence provides a more precise estimation strategy.

2 RELATED WORK

Our work provides a solution to off-policy evaluation of ranking models using implicit feedback [10, 16, 17], that rely on some form of click model [4] to achieve unbiased offline evaluation. The arising bias comes from user behavior such as position bias, trust bias, and selection bias, see for instance [7, 8]. A popular click model is the position-based model, that needs an estimated position bias curve [17]. We make extensive use of Li et al.'s [11] survey of click models for offline evaluation. Other approaches go beyond the full visibility setting such as policy aware approaches [13].

3 BACKGROUND

We are interested in estimating the total number of clicks for a target policy π using recorded interactions from a production policy π_0. The quantity of interest is $\Delta_\pi = \mathbb{E}_x \mathbb{E}_c \mathbb{E}_{Y \sim \pi(\cdot|x)} \left[\sum_{y \in Y} c(y) \right]$, where we compute expectations over context features x, received clicks c and exposed rankings Y that contain items y. We make use of logging data of the form $\mathcal{D} = \left\{ x_i, Y_{0,i}, c(\cdot|Y_{0,i}), \pi_0 \right\}_{i=1}^{n}$ for n different queries, where $Y_{0,i} \sim \pi_0(\cdot|x_i)$ and $c(y|Y_{0,i}) \in \{0, 1\}$ indicates which items $y \in Y_{0,i}$ received a click. Our suggested estimator has the form

$$\hat{\Delta}(\pi|\mathcal{D}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{y \in Y_{0,i}} w \times c(y|Y_{0,i}). \tag{1}$$

The inverse propensity score (IPS) weight w corrects the fact that logging and policy are different and accounts for position biases. We will specify the form of w depending on the underlying assumptions. The expectation of a click on item y is

$$\mathbb{E}_c[c(y|Y)|x] = \begin{cases} \bar{c}(y, \text{rank}(y|Y)|x), & \text{for the item-position model,} \\ \text{rel}(y|x) \times \mathbb{P}(o(y)|\text{rank}(y|Y)), & \text{for the position-based model.} \end{cases} \tag{2}$$

Under the IP model a click depends on item y and its position, defined by its rank, whereas the PB model assumes a click factorizes into relevance $\mathrm{rel}(y)$ and observation probability $\mathbb{P}(o(y)|\mathrm{rank}(y|Y))$ of the item, where $o(y)$ indicates if an item was examined.

Position-based model. The position-based model corrects the fact that not all positions have equal probability of being observed by the user. By weighting clicks using a position bias curve the examination behavior of the user is taken into account. A position bias curve p quantifies the probability of an item being observed in a given position, i.e., $p_k = \mathbb{P}(o(y)|\mathrm{rank}(y|Y) = k)$, where $o(y)$ denotes the event that item y is observed in the position it was displayed in. The corresponding IPS weight (in the full-visibility setting) is

$$w_{pbm}(y|Y, Y_0) = \frac{p_{\mathrm{rank}(y|Y)}}{p_{\mathrm{rank}(y|Y_0)}}. \tag{3}$$

Thus, clicks are weighted according to the visibility ratio of items under logging policy and target policy.

Item-position model. The item-position model [11] does not require a position bias curve. It uses directly the propensities of the logging policy, denoted $\mathbb{P}(\mathrm{rank}(y|Y_0) = \cdot|x, \pi_0)$. The propensities quantify the probability with which an item is displayed in a given rank. The resulting IPS weight is

$$w_{ip}(y|Y, Y_0) = \frac{I\{\mathrm{rank}(y|Y_0) = \mathrm{rank}(y|Y)\}}{\mathbb{P}(\mathrm{rank}(y|Y_0) = \mathrm{rank}(y|Y)|x, \pi_0)}. \tag{4}$$

Here, we assume that the target policy is deterministic, and the logging policy is stochastic. The IP model assigns a weight of 0 if target and logging position mismatch and weights up rewards where target and logging rank agree.

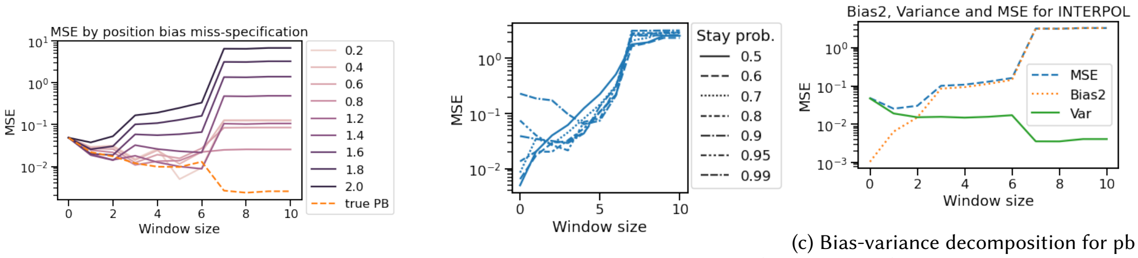
4 INTERPOL ESTIMATOR

The position-based model captures overall user behavior, but it can be biased. The item-position model makes fewer modeling assumptions that may lead to bias, but it can have high variance. We suggest an interpolation between the two approaches. We check if the position where the logging policy ranks an item $\mathrm{rank}(y|Y_0)$ falls inside a window of size T around the position $\mathrm{rank}(y|Y)$ where the target policy ranked it. The size of the window is controlled by the interpolation parameter T, and we denote the event of y being inside the window by $I\{\mathrm{rank}(y|Y_0) \in [\mathrm{rank}(y|Y) \pm T]\}$. Our novel estimator is based on the weight

$$w_T(y|Y, Y_0) = \frac{I\{\mathrm{rank}(y|Y_0) \in [\mathrm{rank}(y|Y) \pm T]\}}{P(\mathrm{rank}(y|Y_0) \in [\mathrm{rank}(y|Y) \pm T]|x, \pi_0)} \times \frac{p_{\mathrm{rank}(y|Y)}}{p_{\mathrm{rank}(y|Y_0)}}. \tag{5}$$

If $T = 0$, then the position of an item has to be identical under the target and logging policy in order to provide a non-zero weight for observed rewards. In this case we recover the IPM as the PBM part is either 1 or ignored in case of a miss-match. For $T > 0$ position bias weights are limited to at most T positions apart. If T is equal to the length of the displayed list $|Y|$ in the full visibility setting, the resulting denominator is equal to 1 as $P(\mathrm{rank}(y|Y_0) \in [\mathrm{rank}(y|Y) \pm T]|x, \pi_0) = 1$ and the PBM is recovered. We denote by $\hat{\Delta}_T(\pi|x, \pi_0, Y_0)$ the estimator that uses the weights as defined in (5) inside the generic estimator in (1).

Proposition 4.1. *$\hat{\Delta}_T(\pi|x, \pi_0, Y_0)$ is an unbiased estimator of Δ_π for all window sizes T if the logging data is generated from a known logging policy π_0 with full support, under the position-based model with a known position bias curve p. (See the appendix for a proof.)*



(a) Stay probability set to 0.95.

(b) Misspecified pb curve set to $p^{1.8}$.

(c) Bias-variance decomposition for pb curve = $p^{1.8}$, stay probability = 0.95.

Fig. 1. Estimated MSE for different position bias curves in Figure 1a. Figure 1b highlights different randomization strengths of the logging policy with the resulting MSE over different window sizes. Figure 1c illustrates a bias-variance decomposition of the MSE for the position bias curve set to $p^{1.8}$ and the stay probability of the logging policy randomization set to 0.95.

5 EXPERIMENTS

We illustrate our off-policy evaluation approach INTERPOL with experiments on a synthetic data set that is simulated in a controlled environment using 5, 000 data points. We highlight the impact of varying (i) the position bias misspecification (using powers of the true curve); (ii) the different window sizes and (iii) the randomization of the logging policy (via random position swaps, which are controlled by a parameter called stay probability) on our offline evaluation. Implementation details are available in the appendix.

Results. When it comes to the interpolation between the IPM (window size 0) and the PBM (window size 10) we see that the arising bias-variance trade-off in Figure 1a is impacted by the misspecification of the position bias curve. As the window size goes up (left to right) the MSE of the estimator first decreases due to a reduction in variance and then increase due to an increase in bias. For all levels of misspecification we eventually end at a clearly biased PBM estimator. For small powers (below 1) the misspecification of the PBM actual acts as weight clipping, which seems to be beneficial in terms of MSE for small window sizes. There seems to be a region around window size 1 to 6, where the MSE is lowest. Consequently, the estimator that offers the best bias-variance trade-off is INTERPOL with a properly chosen window size. Interestingly, the interpolation of INTERPOL can also reduce variance even if the correct position bias curve is used, as highlights Figure 1a, where a large window size of 8 leads to favorable MSE. When varying the randomization of the logging policy, see Figure 1b, we also identify a favorable bias-variance trade-off for a window size between 2 and 6 for a weak randomization of the logging policy (stay probability above 0.9). For stronger randomization the IPM has a lower MSE. Figure 1c illustrates a bias-variance decomposition of the estimator. For a window size larger than 2 the bias starts dominating the MSE.

6 DISCUSSION AND CONCLUSION

We have introduced a novel off-policy estimator, called INTERPOL, for learning-to-rank in the full visibility setting that interpolates the IPM and the PBM. INTERPOL has a favorable MSE, even when the PBM is correctly specified. In future work, we plan to include the top-k setting, investigate the misspecified version of the PBM and study the MSE of different window size T from a theoretical perspective. We also want to extend experiments and provide methods for choosing the best window size, using ideas based on [15]. Finally, we aim to develop off-policy learning approaches based on our interpolation idea.

REFERENCES

[1] Aman Agarwal, Ivan Zaitsev, Xuanhui Wang, Cheng Li, Marc Najork, and Thorsten Joachims. 2019. Estimating position bias without intrusive interventions. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 474–482.

[2] Qingyao Ai, Keping Bi, Cheng Luo, Jiafeng Guo, and W Bruce Croft. 2018. Unbiased learning to rank with unbiased propensity estimation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 385–394.

[3] Walid Bendada, Guillaume Salha, and Théo Bontempelli. 2020. Carousel personalization in music streaming apps with contextual bandits. In *Fourteenth ACM Conference on Recommender Systems*. 420–425.

[4] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. 2015. Click models for web search. *Synthesis lectures on information concepts, retrieval, and services* 7, 3 (2015), 1–115.

[5] Alexandre Gilotte, Clément Calauzènes, Thomas Nedelec, Alexandre Abraham, and Simon Dollé. 2018. Offline a/b testing for recommender systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 198–206.

[6] Rolf Jagerman, Harrie Oosterhuis, and Maarten de Rijke. 2019. To Model or to Intervene: A Comparison of Counterfactual and Online Learning to Rank from User Interactions. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 15–24. https://doi.org/10.1145/3331184.3331269

[7] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2017. Accurately interpreting clickthrough data as implicit feedback. In *Acm Sigir Forum*, Vol. 51. Acm New York, NY, USA, 4–11.

[8] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. 2007. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)* 25, 2 (2007), 7–es.

[9] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *Proceedings of the tenth ACM international conference on web search and data mining*. 781–789.

[10] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) *(WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 781–789. https://doi.org/10.1145/3018661.3018699

[11] Shuai Li, Yasin Abbasi-Yadkori, Branislav Kveton, Shan Muthukrishnan, Vishwa Vinay, and Zheng Wen. 2018. Offline evaluation of ranking policies with click models. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1685–1694.

[12] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In *Proceedings of the 12th ACM conference on recommender systems*. 31–39.

[13] Harrie Oosterhuis and Maarten de Rijke. 2020. *Policy-Aware Unbiased Learning to Rank for Top-k Rankings*. Association for Computing Machinery, New York, NY, USA, 489–498. https://doi.org/10.1145/3397271.3401102

[14] Matteo Ruffini, Vito Bellini, Alexander Buchholz, Giuseppe Di Benedetto, and Yannik Stein. 2022. Modeling Position Bias Ranking for Streaming Media Services. (2022).

[15] Yi Su, Pavithra Srinath, and Akshay Krishnamurthy. 2020. Adaptive estimator selection for off-policy evaluation. In *International Conference on Machine Learning*. PMLR, 9196–9205.

[16] Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miro Dudik, John Langford, Damien Jose, and Imed Zitouni. 2017. Off-policy evaluation for slate recommendation. *Advances in Neural Information Processing Systems* 30 (2017).

[17] Xuanhui Wang, Michael Bendersky, Donald Metzler, and Marc Najork. 2016. Learning to rank with selection bias in personal search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 115–124.

[18] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position bias estimation for unbiased learning to rank in personal search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 610–618.

A THEORETICAL RESULTS

In this section we provide the proof of our main result.

A.1 Unbiasedness of the interpolating estimator under the position-based model

We show that the interpolating estimator is unbiased if user interactions are actually coming from a position-based model, and we dispose of the correct position bias curve under the full visibility setting.

Proof of Proposition 4.1.

Proof. We focus on showing that $\hat{\Delta}_T(\pi|x, \pi_0, Y_0)$ is unbiased for a single sample x. The generalization using the distribution over contexts x is straightforward. We evaluate

$$\mathbb{E}\left[\hat{\Delta}_T(\pi|x, \pi_0, Y_0)\right] \tag{6}$$

$$= \mathbb{E}\left[\sum_{y\in Y_0} \frac{I\{\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]\}}{P(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times \frac{\sum_{k=1}^{K} I\{\text{rank}(y|Y) = k\} \times p_k}{\sum_{k=1}^{K} I\{\text{rank}(y|Y_0) = k\} \times p_k} \times c(y|Y_0)\right], \tag{7}$$

$$= \mathbb{E}_c\mathbb{E}_{\pi_0}\left[\sum_{y\in Y_0} \frac{I\{\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]\}}{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times \frac{p_{\text{rank}(y|Y)}}{p_{\text{rank}(y|Y_0)}} \times c(y|Y_0)\right], \tag{8}$$

$$= \mathbb{E}_{\pi_0}\left[\sum_{y\in Y_0} \frac{I\{\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]\}}{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times \frac{p_{\text{rank}(y|Y)}}{p_{\text{rank}(y|Y_0)}} \times \mathbb{E}_c c(y|Y_0)\right], \tag{9}$$

$$= \mathbb{E}_{\pi_0}\left[\sum_{y\in Y_0} \frac{I\{\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]\}}{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times \frac{p_{\text{rank}(y|Y)}}{p_{\text{rank}(y|Y_0)}} \times \text{rel}(y)\mathbb{P}(o(y)|\text{rank}(y|Y_0))\right], \tag{10}$$

$$= \mathbb{E}_{\pi_0}\left[\sum_{y\in Y_0} \frac{I\{\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]\}}{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times p_{\text{rank}(y|Y)} \times \text{rel}(y)\right], \tag{11}$$

$$= \sum_{y\in Y_0} \frac{\mathbb{E}_{\pi_0}\left[I\{\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]\}\right]}{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times p_{\text{rank}(y|Y)} \times \text{rel}(y), \tag{12}$$

$$= \sum_{y\in Y_0} \frac{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)}{\mathbb{P}(\text{rank}(y|Y_0) \in [\text{rank}(y|Y) \pm T]|x, \pi_0)} \times p_{\text{rank}(y|Y)} \times \text{rel}(y), \tag{13}$$

$$= \sum_{y\in Y_0} p_{\text{rank}(y|Y)} \times \text{rel}(y) = \mathbb{E}_c\mathbb{E}_\pi\left[\sum_{y\in Y} c(y)\right]. \tag{14}$$

Line (7) is obtained by using the definition of (1). Line (8) uses the definition of the position bias weights and decomposes the expectation. Line (9) pulls the expectation of the click model inside the sum, exploiting the linearity of the expectation and the fact that clicks do not interact by assumption of the PBM. Line (10) uses the definition of a click under the PBM and in line (11) we simplify the production of examination probability. In Line (12) we pull the expectation with respect to the logging policy inside the sum and line (13) evaluates this expectation. Finally line (14) simplifies the expression and uses the previous identities in reverse using the position-based model under the target policy π. □

B EXPERIMENTS

We describe our experiment in more detail and provide more results on the estimator in the full visibility setting.

Data generation for the toy experiment Our synthetic data generation uses a toy model that allows to easily compute expectations of the true reward as well as to control the strength of logging policy randomization and the level of position bias misspecification. This experimental set-up is not supposed to be realistic, but rather to study properties of INTERPOL easily. We generate $5,000$ observations from a synthetic logging policy that ranks $K = 10$ different actions. The true position bias curve is given as $p = [1, 0.9, 0.8, \cdots, 0.1]$, the biased curve is defined as p^x component wise, where $x \in [0.2, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2.0]$. We illustrate the curves for p^x in Figure 2. The relevant items are items $[1, 2, 4, 7]$ and the logging policy π_0 orders items $[6, 0, 3, 1, 4]$ at the top and items $[7, 5, 2]$ at the bottom of the list. The other items are displayed in arbitrary order. Additionally, the logging policy swaps the ranked items randomly, where every item has a probability of $q\%$ of staying in its original position and a probability of $(100 - q\%)/9$ of being ranked in all other positions. We set these probabilities to $[50\%, 55\%, 60\%, 70\%, 80\%, 90\%, 95\%, 99\%]$. We denote this probability *stay probability*. The target policy π deterministically ranks items $[7, 0, 3, 1]$ at the top and items $[2, 4]$ at the bottom. A positive reward of 1 is generated for the relevant items and this reward is revealed according to the examination probability (i.e., position bias curve p). Hence, the expected reward for the target policy (under full visibility) is $1 \times p_0 + 1 \times p_3 + 1 \times p_8 + 1 \times p_9 = 2$, where $p_0 = 1, p_3 = 0.7, p_8 = 0.2, p_9 = 0.1$.

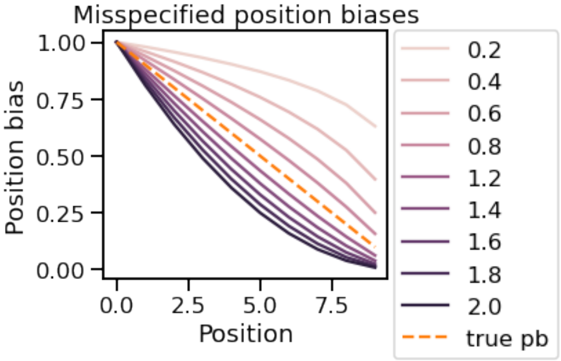


Fig. 2. Position bias curves used in the experiments.

Figure 3 highlights the behavior of the item-position model and the position-based model. The left-hand figure uses a misspecified position bias curve and clearly the resulting estimate is severely biased. The IP model (middle figure) exhibits higher variance but the true reward (straight line) lies inside the 95 % confidence interval of the IPM estimator. As the data set size increase, the estimator gets more precise and the confidence intervals shrink around the true value (2.0). The PBM (right-hand figure) that uses the correct position bias curve estimates the reward correctly and has less variability than the IP model. For illustrative purpose we also show the interpolation over different window sizes in Figure 4.

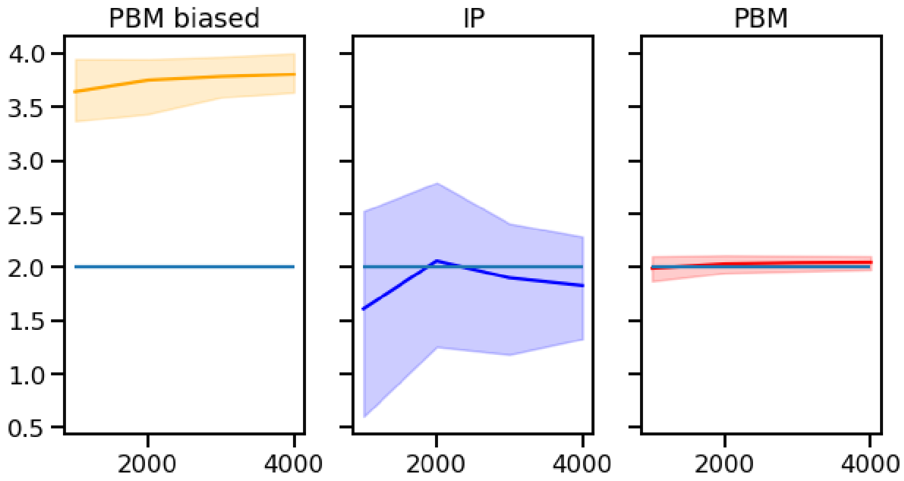


Fig. 3. Estimated reward for different estimators over different data set size in the full visibility setting. Stay probability set to 95% and the misspecified position bias curve is $p^{1.8}$.

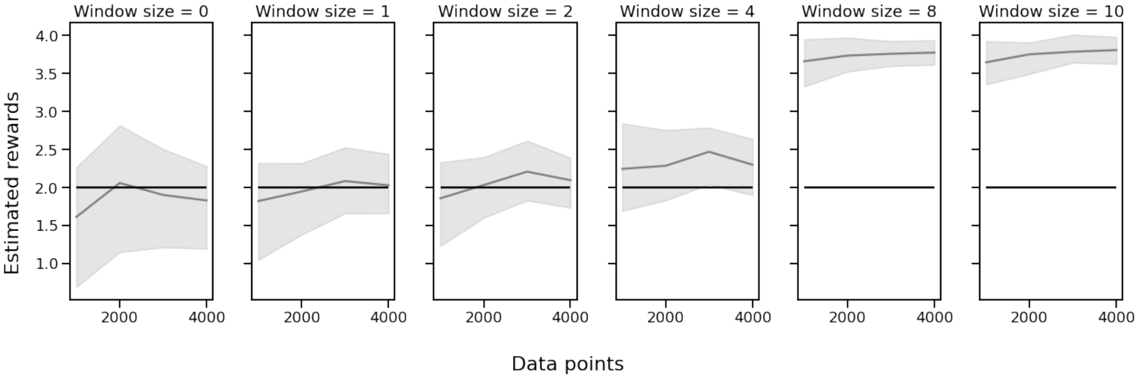


Fig. 4. Interpolation of INTERPOL in the full visibility setting. Stay probability set to 95% and the misspecified position bias curve is $p^{1.8}$.