A/B Testing Adaptations based on Possibilistic Reward Methods for Checkout Processes: A Numerical Analysis

Miguel Martín, Antonio Jiménez-Martín[a] and Alfonso Mateos[b]

Decision Analysis and Statistics Group, Universidad Politécnica de Madrid, Campus de Montegancedo S/N,
Boadilla del Monte, 28660, Madrid, Spain

Abstract: A/B Testing can be used in digital contexts to optimize the e-commerce purchasing process so as to reduce customer effort during online purchasing and assure that the largest possible number of customers place their order. In this paper we focus on the checkout process. Most of the companies are very interested in agilice this process in order to reduce the customer abandon rate during the purchase sequence and to increase the customer satisfaction. In this paper, we use an adaptation of A/B testing based on multi-armed bandit algorithms, which also includes the definition of alternative stopping criteria. In real contexts, where the family to which the reward distribution belongs is unknown, the possibilistic reward (PR) methods become a powerful alternative. In PR methods, the probability distribution of the expected rewards is approximately modeled and only the minimum and maximum reward bounds have to be known. A comparative numerical analysis based on the simulation of real checkout process scenarios is used to analyze the performance of the proposed A/B testing adaptations in non-Bernoulli environments. The conclusion is that the PR3 method can be efficiently used in such environments in combination with the PR3-based stopping criteria.

1 INTRODUCTION

In the current market of digital services an content (retail, media, news, e-commerce) there is a continue necessity in offering the best user experience by providing the customer with the right content that they most likely to use and by offering and effortless access to any task, transaction and process required to complete a service.

The most used approach for this and other types of service or user interface optimization solutions is to continuously make changes to the offered services or interfaces and use a specific indicator to measure which change produces the best expected indicator value. This type of experimentation is commonly known as *A/B testing*.

In recent years, some companies and solutions (for instance, Google Analytics) have addressed this experimentation process as a multi-armed bandit (MAB) problem (Audibert and Bubeck, 2010; Baransi et al., 2014; Chapelle and Li, 2011; Garivier and Cappé, 2011; Kaufmann et al., 2012; Martín

[a] https://orcid.org/0000-0002-4947-8430
[b] https://orcid.org/0000-0003-4764-6047

et al., 2018), using algorithms existing in the literature designed to optimize the conflict between exploring all the variations and exploiting the best variation. This drastically reduces the number of unnecessary experiments. This is known in the A/B testing market as *dynamic traffic distribution*.

The most used MAB algorithm is Thompson sampling (Chapelle and Li, 2011), since it performs well under delayed rewards conditions typical of A/B testing. However, Thompson sampling can only be used if the type of distribution associated with the rewards or the indicator to be optimized is known a priori. Bernoulli distributions often used to measure the success or failure of an action (whether or not the customer makes the purchase or views a content). In other cases, however, the distribution type may be unknown and depend on factors such as purchase price, navigation time or number of pages visited before purchase.

A complementary technique used to optimize the performance of the A/B Testing is to improve the experiment stopping criterion. To do this, some solutions perform the hypothesis tests using a Bayesian approach to ascertain the statistical significance. As in Thompson sampling, however, the distribution

family to which the rewards belong has to be known a priori in order to perform this type of Bayesian analysis.

This drawback is the main reason why most of the AB Testing using MAB algorithms and new stopping criteria are limited most of the case to Bernoulli experiments, however, the numerical analyses carried out recently in (Martín et al., 2019) and (Martín et al., 2020) show that the *possibilistic reward* (PR) methods (Martín et al., 2018) outperform other MAB algorithms in scenarios with delayed rewards and also where the associated reward distribution does not have to be known: Test A/B in digital content web when the reward is continue and increase if customer read the content and campaign management in digital marketing recommendation systems. PR methods approximate a distribution function for rewards that can also be used to perform classic A/B testing, albeit with a stopping condition based on Bayesian and non-frequentist hypothesis tests.

One of the most common business cases where is continue optimizing scenarios where the reward distribution is unknown are those related with the purchasing process so as to reduce customer effort to complete the online purchasing process and assure that the largest possible number of customers place their order in e-commerce companies

In this paper we extend experiments carried out (Martín et al., 2020) (using the adaptation A/B testing to account for PR methods, together with the definition of a new stopping criterion also based on PR methods to be used for both classical A/B testing and A/B testing based on MAB algorithms) with a common scenarios of optimize purchasing process in e-commerce companies.

The paper is structured as follows. Section 2 briefly reviews possibilistic reward (PR) methods. Section 3 gives a brief description of A/B testing and improvements aimed at optimizing how tests are carried out (dynamic traffic distribution and stopping criterion) and the use of PR methods in A/B testing. Section 4 describes the numerical analysis carried out for checkout process scenarios and the results. Finally, some conclusions and future research work are outlined in Section 5.

2 POSSIBILISTIC REWARD METHOD

Possibilistic reward methods (*PR1,PR2* and *PR3*) (Martín et al., 2017; Martín et al., 2018) have recently been proposed as an alternative to MAB algorithms in

the literature. A review of the most important allocation strategies can be found in (Martín et al., 2017).

The basic idea of the *PR1* method is as follows: the uncertainty about the arm expected rewards are first modelled by means of possibilistic reward distributions derived from a set of infinite nested confidence intervals around the expected value on the basis of the Chernoff-Hoeffding inequality (Hoeffding, 1963).

Then, the method follows the *pignistic probability transformation* from decision theory and the transferable belief model (Smets, 2000). The pignistic probability transformation establishes that when we have a plausibility function, such as a possibility function, and any further decision-making information, we can convert this function into an probability distribution following the *insufficient reason principle*.

Once we have a probability distribution for the reward of each arm, then a simulation experiment is carried out by sampling from each arm according to their probability distributions to find out which one has the highest expected reward. Finally, the selected arm is played and a real reward is output.

As mentioned above, the PR1 method starting point is the Chernoff-Hoeffding inequality (Hoeffding, 1963). This inequality establishes an upper bound on the probability that the sum of random variables deviates from its expected value for [0,1], which can be used to build an infinite set of nested confidence intervals.

The difference between PR1, PR2 and PR3 lies in the type of concentration applied and subsequent approximations. PR1 and PR2 are based on the Hoeffding concentration, whereas PR3 is based on a combination of the Chernoff and Bernstein concentrations.

A numerical study based on of five complex and representative scenarios was performed in (Martín et al., 2018) to compare the performance of PR methods against other MAB methods in the literature. *PR2* and *PR3* methods perform well in all representative scenarios under consideration, and are the best allocation strategies if truncated Poisson or exponential distributions in [0,10] are considered for the arms. Besides, Thompson sampling (TS), PR2 and PR3 perform equally with a Bernoulli distribution for the arm rewards. PR2 is exactly the same as the generalization of the TS method proposed in (Agrawal and Goyal, 2012) (see Algorithm 2).

Moreover, the numerical analyses conducted recently in (Martín et al., 2019) show that *possibilistic reward* (PR) methods outperform other MAB algorithms in digital marketing content recommendation systems for campaign management, another scenario with delayed rewards.

Finally, PR methods have one big advantage over other MAB algorithms, including TS: all they need to know is the interval to which the reward belongs rather than the total reward distribution. PR methods approximate a distribution function for rewards that can also be used to perform a classic A/B test, albeit with a stopping condition based on Bayesian and non-frequentist hypothesis tests. In this way, experimentation can be efficiently carried out with these methods in contexts where the objective is not confined merely to action success or failure (Bernoulli distribution) but also to the minimization of the total number of page views or the duration of a session, or the maximization of the total income from web e-commerce.

3 A/B TESTING IN DIGITAL SERVICES

It is common practice in companies that offer services and products through online channels (web and mobile apps) to continuously optimize their user interfaces with the aim of improving one or more of their key business indicators, such as customer satisfaction, online sales, content consumption times, or advertising conversion rates.

These experiments are known in the industry as *A/B testing*, randomized control trials (RCT) where different variations are tested until there is statistical significance.

Two options are currently used to implement A/B testing:

- *Ad hoc* developments, mainly using proprietary software (primarily large content managers, such as Google, Facebook, Netflix, Amazon...), or libraries, such as Facebook Planout, and plug-ins by e-commerce platforms, such as Magento or Pentashop.

- Specialized experimentation software, where there is a wide variety of vendors, notably Google Optimizer, Optimizely, AB Tasty and VWO.

The most advanced experimentation or A/B testing solutions have incorporated improvements aimed at optimizing how the tests are carried out. This optimization consists of achieving statistical significance with the lowest opportunity cost, that is, experiment with the worst variations as few times as possible, since they result in worse performance than the best variation. To do this, two optimization processes account for dynamic traffic distribution and the stopping criterion, respectively.

3.1 Dynamic Traffic Allocation

In A/B testing, traffic is originally distributed equally for each of the variations to provide the same number of experiments. However, it is more efficient to redistribute traffic dynamically, sending more or fewer experiments to variations perform better or worse, respectively, provided that statistical significance is achieved.

Traffic can be distributed dynamically using multi-armed-bandit (MAB) methods. In this context, the decisions have to be taken each time a user/customer accesses the web page by selecting the variation (arm) that will be shown to that user/customer. Then, a delayed stochastic reward will be received depending on the scenario under consideration (Bernoulli or other reward distributions). The aim is then to select a strategy (a sequence of variation selections) that optimizes the expected reward value, i.e. minimizing the expected regret or the opportunity cost.

The three advanced solutions for optimizing the dynamic traffic distribution using MAB methods for an objective with a Bernoulli distribution (success or failure), such as conversion ratios, click ratios, are:

1. A Thompson sampling variation for Bernoulli rewards. In this case, the original version is *probability matching*, where a weight consisting of the probability of its expected value being better than the rest is computed for each variation. The variation to be executed is then chosen based on a random sample where the probability of selecting each variation corresponds to the previously computed weight. In this way, the variations that are more likely to be the best will be chosen more times.

   This is the dynamic traffic distribution technique based on MAB methods most used by vendors, since it performs very well, even under delayed rewards conditions (Martín et al., 2019), where there is a time delay between a variation and its feedback, as in commercial A/B testing systems.

2. A variation of e-greedy algorithms. Some vendors opt for this simpler algorithm, although convergence is linear rather than logarithmic. According to this approach, the best variations are calculated periodically or after a series of N iterations. Then 80% of the traffic is uniformly distributed to the best variations in order to optimize the expected value and the other 20% is distributed uniformly to all variations in order to perform exploration tasks until the next iteration.

   The companies adopting this approach include Adobe, with Adobe Target.

3. It is quite plausible, although it has not, as far as we know, been published, that some large companies developing their own *ad hoc* experimental software, use Thompson sampling in its most efficient form in this context, where the algorithm is dynamically updated at each decision and not every *N* decisions.

In experiments where the objective follows a non-Bernoulli distribution, measuring browsing time, number of pages visited, total revenue, Thompson sampling cannot be used since it is not possible to parameterize the reward distribution. Therefore, the main companies, such as Adobe and ABTasty, use other alternatives, mainly a variation on e-greedy algorithms. Other vendors such as Google, with Google Analytics, and ABTasty, do not provide information on whether or not and how they perform dynamic traffic distribution with objectives not following a Bernoulli distribution.

As cited before, the type of distribution associated with the rewards or the indicator to be optimized does not have to be known a priori in the possibilistic reward (PR) methods. Thus, they constitute an alternative for dynamic traffic distribution for non-Bernoulli reward distributions.

In (Martín et al., 2020) a variation for dynamic traffic distribution in A/B testing accounting for PR methods for non-Bernoulli reward distributions is proposed.

3.2 Stopping Criterion

The stopping criterion plays a key role in the execution of A/B testing experiments. It is used to decide when a variation is considered to be the best.

The *de facto* method used to define the stopping criterion in most approaches is based on a classical hypothesis test. However, classic stopping criteria are not very efficient, since they are unable to dynamically stop the test when there is enough evidence to suggest that one variation is better than the others (Scott, 2015).

Recently, the most innovative companies are introducing more dynamic stopping criteria to reduce testing costs, leading to the same statistical significance in a similar way. These new methods, although perfectly applicable to classical A/B testing, come hand in hand with the new methods for dynamic traffic distribution. The multi-armed bandit paradigm is the most popular, since the number of samples that have to be executed for each variation is determined dynamically rather than using classical hypothesis tests to identify the number of samples required to achieve statistical significance.

These new criteria are based on different approaches (Bayesian, inequalities bounds...). In (Martín et al., 2020) a review of the most important approaches is provided, including Google Analytics, which uses a stopping criterion based on a Bayesian approach (Scott, 2015; Google,), and Adobe Target (Adobe,), in which a stopping method based on confidence intervals computed by the Bernstein inequality (Bernstein, 1946) is used. Google Analytics and Adobe Target are the the stopping criteria most used by the main vendors.

The stopping method based on the *value remaining* used by Google Analytics (Scott, 2015) is very efficient in environments with rewards following a Bernoulli distribution, since it has to know the exact distribution of the expected rewards in order to carry out the simulations. The distribution of the expected rewards is inferred with a Bayesian approach.

This approach, however, has a drawback: the shape of the reward distribution has to be known or modeled by a family of parameterizable distributions on which priors can be applied. In addition, it should be tractable or at least computationally efficient to update the a posteriori distributions and the expected value. This is not very often the case in many real contexts, where the family to which the reward distribution belongs (normal, Poisson, Bernoulli) is unknown. Besides, if the distribution is known or can be modeled, it is very difficult to make an efficient inference using, for example, conjugate priors.

To overcome this problem, a new approach was proposed in (Martín et al., 2020), in which the probability distribution of the expected rewards efficiently is approximately modeled by applying the possibilistic rewards methods (PR2 and PR3) for the reward in each variation. To do this, only the minimum and maximum reward bounds have to be known rather than the distribution of each reward. This information is commonly available in real contexts.

Once the density function of the expected reward (Step 3 in PR2 and PR3) is derived, the simulation and stopping condition techniques used in (Scott, 2015) are applied. In the Section 5, reporting a numerical analysis of these methods on checkout process scenarios, these approaches will be denoted as *PR2 ValRem* and *PR3 ValRem*.

Besides, a stopping criterion computed from approximations to the probability distributions of the expected reward derived from PR2 and PR3 methods is also proposed in (Martín et al., 2020) for emulating confidence level-based stopping criteria, such as empirical Bernstein in Adobe Target.

To do this, function that outputs the percentile value is needed, which will be used as a confidence

level, from distributions PR2 or PR3. As PR2 and PR3 are Beta distributions, we can use the quantile function, also called *ppf* (percentile point function), to compute these dimensions. This function can be analytically obtained and is available in any statistical software library.

Once these dimensions have been derived, we have practically the same stopping criterion as the one used by Adobe Target.

In the Section 5, these approaches will be denoted as *PR2 bounds* and *PR3 bounds*.

4 NUMERICAL EXPERIMENTS AND RESULTS

In most e-commerce companies the purchase process, also called *checkout process* in business argot, starts when a customer after searching and evaluating some service products he/she is interested in, decides to buy one or several. This process usually start by clicking the buy button associated to the product in the website, or by clicking the checkout button and go to buy the products previously added to the cart.

The checkout process in most cases consists of the following tasks:

- Confirming from the cart the products and quantities the customer wants to purchase.

- Selecting a pay method (card, PayPal, etc.) and providing the payment data (credit card number, expiration date, etc.)

- Logistic information regarding transport duration and fees is displayed, and the address information is entered by the customer.

- The customer is sometimes asked to sign-in or sign-up and/or offered a coupon, some cross-selling or up-selling products.

- Finally, the customer is requested to confirm all the entered information to process the purchase, and the corresponding recipe order is displayed.

The different variations of the process usually consist of grouping or splitting the different tasks in different steps, adding and/or removing steps, in order to check what variations and designs are more efficient.

In this optimization process, reducing the number of abandons becomes crucial but also the time the customer spends in the purchase. Therefore, the objetive or reward will be a time function, in which the reward is 0 if the customer abandons the process and, otherwise the reward is higher the less time is spent in the purchase.

The checkout process consists on 1 to *n* steps, corresponding to pages necessary to read or enter some information. In any step, we will simulate an abandon rate by means of a Bernoulli distribution. If the customer abandons the process then we have a reward 0, otherwise the user has spent some time in this step and go to next step. A gamma distribution is used to generate the times spent in the different steps and the total time of customer purchase process will be the sum of the times spent in the steps throughout the checkout process.

We have simulated two different scenarios with only one and with more than one variations.

4.1 Checkout Process Scenario with Only One Variation

In this scenario the current state compared against only one variation. The current process consists of a purchase process with two steps: A step to enter all the information (pay methods, pay data, name, address, etc.) and a second step to review information and confirm the purchase.

The first step has an abandon rate of 10% and the time spent in this step follows a Gamma distribution with parameter $\alpha = 540$ and $\beta = 0.16$, what leads to a mean time of 85 seconds. The abandon rate in the second step is 1% and the time spent in this step follows a Gamma distribution with parameter $\alpha = 140$ and $\beta = 0.16$, with a mean time of 21 seconds.

Finally, the maximum timeout to complete the purchase is 300 seconds.

The challenger checkout process we will have the next configuration. It consists of three steps: A step to enter all the payment information (pay methods, pay data, name), a second one to enter logistic information and a third step to review information and confirm the purchase.

The first step has an abandon rate of 5% and the time spent in this step follows a Gamma distribution with parameter $\alpha = 340$ and $\beta = 0.16$, with a mean time of 55 seconds. The abandon rate in the second step is 5% and the time spent in this step follows a Gamma distribution with parameter $\alpha = 240$ and $\beta = 0.16$, with a mean time of 39 seconds. Finally, in the third the gamma parameters are $\alpha = 140$ and $\beta = 0.16$, with a mean time of 21 seconds.

The maximum timeout to complete the purchase also is 300 seconds.

The reward distributions for the variations are unknown for the tested algorithms and the aim is to analyse throughout the simulation their reinforcement learning capacity on the basis of the accumulated regrets together with the number of samples necessary

for the corresponding stopping criterion.

Table 1 shows the Scenario 1 results for all the combinations of methods and stopping criteria under consideration. Mean values are provided in all columns derived from 500 simulations, and the methods are ordered from lowest to highest mean accumulated regret. Besides, the accumulated regret density for the best 10 combinations are shown in Fig. 1.

Table 1: Results in Scenario 1 with only one variation.

| method	stopping crit.	accum. regret	std. dev.	samples
PR3	PR3 ValRem	11.394	7.952	2604.25
PR2	PR3 ValRem	12.366	7.610	2014.75
PR3	PR3 bounds	16.284	9.574	6655.90
A/B test	PR3 ValRem	16.398	9.318	1917.25
e-Greedy	PR3 ValRem	19.298	3.954	2252.75
PR2	PR3 bounds	20.694	10.031	4387.10
PR3	PR2 ValRem	21.407	11.237	19912.50
PR2	PR2 ValRem	26.623	13.537	6727.75
e-Greedy	PR3 bounds	28.746	12.341	3358.80
A/B test	PR3 bounds	31.501	13.803	3691.80
PR3	PR2 bounds	33.708	11.698	309003.85
e-Greedy	PR2 ValRem	38.334	18.029	4482.25
A/B test	PR2 ValRem	43.608	18.840	5107.25
PR2	PR2 bounds	51.540	19.331	30505.50
e-Greedy	PR2 bounds	117.740	34.735	13778.80
A/B test	PR2 bounds	117.756	33.241	13781.40
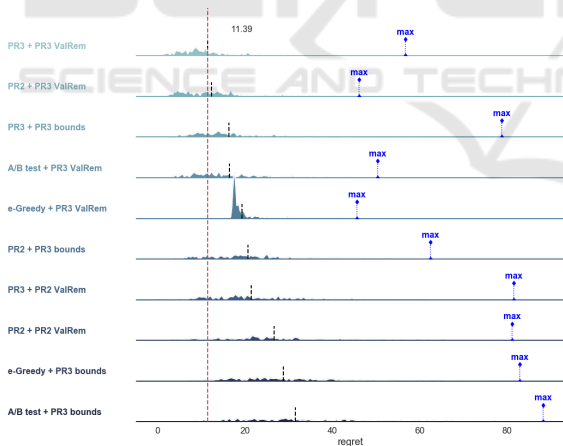
Figure 1: Ripple graph of Scenario 1.

First is important to point out that the mean accumulated regret derived from the classical A/B testing with also the classical stopping criterion is 248.41, whereas the number of samples needed is 28228, being both measures clearly outperformed by the combinations analysed in the numerical study.

In this scenario, PR3 + PR3 ValRem is the best combination, followed by PR2 + PR3 ValRem and PR3 + PR3 bounds in terms of mean accumulated regrets. Moreover, the three best-ranked combinations

in terms of mean accumulated regrets also good values with respect to standard deviations (dispersion of accumulated regrets). They are only outperformed by the combination e-Greedy +R3 ValRem, but with a worst mean accumulated regret. However, PR2 + PR3 ValRem slightly outperforms the other two combinations in terms of maximum accumulated regret (see Fig. 1).

We can also find that for the same stopping criterion, PR3 is always better than the rest of the algorithms: The accumulated mean regret and standard deviation (except for the combination e-Greedy + PR3 ValRem with a lower std. deviations) are lower.

Regarding the stopping criteria, we find, looking at the mean samples column, that the values for combinations with the PR3 value remaining (PR3 ValRem) stopping criterion are the lowest, followed by PR2 ValRem and PR3 bounds (depending on the combination) and PR2 bounds. The two best-ranked combinations in terms of mean accumulated regrets are among the best with respect to the number of mean required samples.

We can conclude that the combinations PR3 + PR3 ValRem and PR2 + PR3 ValRem outperforms the other combinations in this Scenario 1 with good performances regarding the mean samples.

4.2 Checkout Process Scenario with 7 Variations

In this scenario 7 possible variations are consider to test the behaviour of algorithms under consideration. First, we have considered 5 variation types as follows:

Variation types 1, 2 and 3 consist of a purchase process with two steps: A step to enter all the information (pay methods, pay data, name, address, etc.) and a second step to review information and confirm the purchase. Variation types 4 and 5 consist of three steps: A step to enter all the payment information (pay methods, pay data, name), a second one to enter logistic information and a third step to review information and confirm the purchase.

The times spent in each of the different steps in the 5 variations under consideration follow a Gamma distribution with parameter $\beta = 0.16$ and the values for parameter α shown in Table 2, which also includes the corresponding mean time spent and abandon rate for each step and variation.

Fig. 2 shows the reward distribution for the five variation types under consideration. Note the different heights in the vertical bar for the reward 0 corresponding to 5 variation types under consideration, which match up with the corresponding abandon rates for each variation type in Table 2. We can see that the

best variation types are the 1 and 4, which also match up with the mean times included in Table 2 since although the mean accumulated time spent in their steps are lower than in the other variations, the accumulated abandon rates are lower (see the vertical lines associated to the zero reward value in the figure).

The above reward distributions for the variations are unknown for the tested algorithms and the aim is to analyse throughout the simulation their reinforcement learning capacity on the basis of the accumulated regrets together with the number of samples necessary on the basis of the stopping criteria.

Finally, in the five variations the maximum timeout to complete the purchase also is 300 seconds.

In the simulation process carried out we consider a variation of types 1 and 2 and two variations of types 3, 4 and 5 each.

Table 2: Variation parameters ($\beta = 0.16$) in Scenario 2.

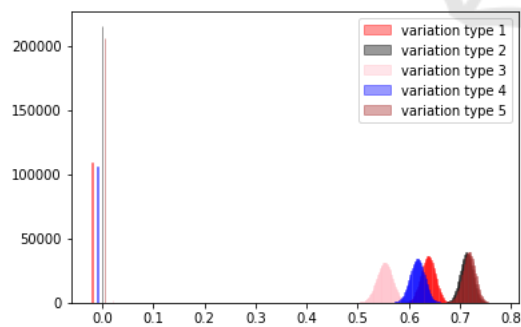| Variation	Step	α	mean time	abandon rate
Variation type 1	step1	540	85 sec.	10%
	step2	140	21 sec.	1%
Variation type 2	step1	440	70 sec.	20%
	step2	100	15 sec.	2%
Variation type 3	step1	640	110 sec.	0.2%
	step2	200	30 sec.	0.02%
Variation type 4	step1	340	55 sec.	5%
	step2	240	39 sec.	5%
	step3	140	21 sec.	1%
Variation type 5	step1	240	39 sec.	10%
	step2	200	30 sec.	10%
	step3	100	15 sec.	2%

Figure 2: Reward distribution in Scenario 2.

Table 3 shows the Scenario 2 results for all the combinations of methods and stopping criteria under consideration. Mean values are provided in all columns derived from 500 simulations, and the methods are ordered from lowest to highest mean accumulated regret. Besides, the accumulated regret density for the combinations are shown in Fig. 3.

Is important to point out again that the mean accumulated regret derived from the classical A/B testing

with also the classical stopping criterion is 2819.334, whereas the number of samples needed is 211980, being both measures clearly outperformed by most of the combinations analysed in the numerical study.

In this scenario, PR3 + PR3 bounds is the best combination, followed by PR3 + PR3 ValRem and PR2 + PR3 ValRem in terms of mean accumulated regrets. Moreover, the three best-ranked combinations in terms of mean accumulated regrets also the best values with respect to standard deviations (dispersion of accumulated regrets) and in terms of maximum accumulated regret (see Fig. 3).

Regarding the stopping criteria, first we can see that the mean number of samples for the combinations in this Scenario 2 are much higher than in Scenario 1 since a more variations are considered in Scenario 2. We find, looking at the mean samples column (last column), that the values for combinations with a PR3-based stopping criterion are the lowest. The three best-ranked combinations in terms of mean accumulated regrets are among the best with respect to the number of mean required samples.

We can conclude that the combinations PR3 + PR3 ValRem and PR2 + PR3 ValRem outperforms the other combinations in this Scenario 1 with good performances regarding the mean samples.

Table 3: Results in scenario 2 with seven variations.

method	stopping crit.	acc. regret	std. dev.	samples
PR3	PR3 bounds	227.55	64.05	5.179e+04
PR3	PR3 ValRem	243.51	70.22	5.718e+04
PR2	PR3 ValRem	259.64	78.12	3.714e+04
PR2	PR3 bounds	364.11	124.84	5.830e+04
A/B test	PR3 ValRem	599.94	267.34	5.522e+04
e-Greedy	PR3 ValRem	609.77	223.52	6.317e+04
PR2	PR2 ValRem	609.92	142.70	1.209e+05
PR2	PR2 bounds	849.74	212.81	2.052e+05
e-Greedy	PR3 bounds	853.88	356.47	8.947e+04
PR3	PR2 ValRem	1182.72	341.18	4.978e+06
PR3	PR2 bounds	1183.33	340.75	4.999e+06
A/B test	PR3 bounds	1190.37	571.33	1.095e+05
e-Greedy	PR2 ValRem	1654.84	528.65	1.760e+05
A/B test	PR2 ValRem	2051.88	781.56	1.888e+05
e-Greedy	PR2 bounds	3182.26	985.21	3.426e+05
A/B test	PR2 bounds	5101.73	1456.92	4.696e+05

5 CONCLUSIONS

In this paper we analyze the use of the A/B Testing to optimize the e-commerce purchasing process, specifically the checkout process, aimed at reducing the customer abandon rate during the purchase sequence and to increase the customer satisfaction reducing the time
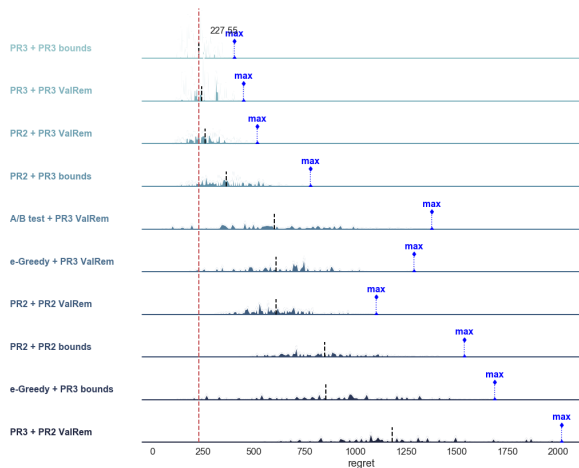
Figure 3: Ripple graph of Scenario 2.

required to end it.

A numerical study has been carried out to compare different adaptations of the A/B Testing based on multi-armed bandit algorithms, also including alternative stopping criteria.

First, we can conclude the the different adaptations of the A/B Testing on the basic of possibility reward (PR) methods together with the alternative stopping criteria outperform the classical A/B Testing in terms of both the mean accumulated regret and the number of samples necessary.

The PR3-based adaptation is the best one in the two scenarios under consideration, together with the PR3-based stopping criteria (PR3 ValRem and PR3 bounds). PR3-based adaptation outperforms the PR2-based adaptation because it better takes advantage of the sample variance to limit the distribution function of the regret expected value, then becoming much better than the PR2-based adaptation the lower is the reward variance.

These conclusions match up with those reached in (Martín et al., 2020), in which a comparative numerical analysis based on the simulation of real scenarios is used to analyze the performance of the same A/B Testing adaptations in both Bernoulli and non-Bernoulli environments.

ACKNOWLEDGEMENTS

REFERENCES

Adobe. Adobe target automatic traffic allocation. https://docs.adobe.com/content/help/en/target/using/activities/auto-allocate/automated-traffic-allocation.html.

Agrawal, S. and Goyal, N. (2012). Analysis of thompson sampling for the multi-armed bandit problem. In *Proceedings of the 25th Anual Conference on Learning Theory*, pages 39.1– 39.26.

Audibert, J.-Y. and Bubeck, S. (2010). Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 11:2785–2836.

Baransi, A., Maillard, O., and Mannor, S. (2014). Sub-sampling for multi-armed bandits. In *Proceedings of the European Conference on Machine Learning*, page 13.

Bernstein, S. (1946). *Probability Theory*. GTTI, Moscow-Leningrad.

Chapelle, O. and Li, L. (2011). An empirical evaluation of thompson sampling. *Advances in Neural Information Processing Systems*, 17:2249–2257.

Garivier, A. and Cappé, O. (2011). The kl-ucb algorithm for bounded stochastic bandits and beyond. *arXiv*.

Google. Google analytics help. https://support.google.com/analytics/answer/2846882/.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Advances Applied Mathematics*, 58:13–30.

Kaufmann, E., Cappé, O., and Garivier, A. (2012). On bayesian upper confidence bounds for bandit problems. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 592–600.

Martín, M., Jiménez-Martín, A., and Mateos, A. (2017). Possibilistic reward method for the multi-armed bandit problem. In *Proceedings of the 6th International Conference on Operations Research and Enterprise Systems*, pages 75–84.

Martín, M., Jiménez-Martín, A., and Mateos, A. (2018). Possibilistic reward methods for the multi-armed bandit problem. *Neurocomputing*, 310:201–212.

Martín, M., Jiménez-Martín, A., and Mateos, A. (2019). A numerical analysis of allocation strategies for the multi armed bandit problem under delayed rewards conditions in digital campaign management. *Neurocomputing*, 363:99–113.

Martín, M., Jiménez-Martín, A., and Mateos, A. (2020). Improving a/b testing on the basis of possibilistic reward methods: a numerical analysis. *Journal of Machine Learning Research*, under review.

Scott, S. L. (2015). Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry*, 31(1):37–45.

Smets, P. (2000). Data fusion in the transferable belief model. In *Proceedings of the 3rd International Conference on Information Fusion*, pages 21–33.