Dynamic pricing using Thompson Sampling with fuzzy events

Citation for published version (APA):
Rhuggenaath, J., de O. da Costa, P. R., Zhang, Y., Akcay, A., & Kaymak, U. (2020). Dynamic pricing using Thompson Sampling with fuzzy events. In M-J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, & R. R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems - 18th International Conference, IPMU 2020, Proceedings (pp. 653-666). (Communications in Computer and Information Science; Vol. 1237 CCIS). https://doi.org/10.1007/978-3-030-50146-4_48

DOI:
10.1007/978-3-030-50146-4_48

Document status and date:
Published: 15/06/2020

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 22. Aug. 2021
Dynamic pricing using Thompson Sampling with fuzzy events

Jason Rhuggenaath, Paulo Roberto de Oliveira da Costa, Yingqian Zhang, Alp Akcay, and Uzay Kaymak

Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands
{j.s.rhuggenaath, p.r.d.oliveira.da.costa, yqzhang, a.e.akcay, u.kaymak}@tue.nl

Abstract. In this paper we study a repeated posted-price auction between a single seller and a single buyer that interact for a finite number of periods or rounds. In each round, the seller offers the same item for sale to the buyer. The seller announces a price and the buyer can decide to buy the item at the announced price or the buyer can decide not to buy the item. In this paper we study the problem from the perspective of the buyer who only gets to observe a stochastic measurement of the valuation of the item after he buys the item. Furthermore, in our model the buyer uses fuzzy sets to describe his satisfaction with the observed valuations and he uses fuzzy sets to describe his dissatisfaction with the observed price. In our problem, the buyer makes decisions based on the probability of a fuzzy event. His decision to buy or not depends on whether the satisfaction from having a high enough valuation for the item out weights the dissatisfaction of the quoted price. We propose an algorithm based on Thompson Sampling and demonstrate that it performs well using numerical experiments.

Keywords: dynamic pricing · Bayesian modeling · exploration-exploitation trade-off · probability of fuzzy events

1 Introduction

In this paper we study a repeated posted-price auction [17] between a single seller and a single buyer. In a repeated posted-price auction there is a single seller and a single buyer that interact for a finite number of periods or rounds. In each round, the seller offers the same item for sale to the buyer. The seller announces a price and the buyer can decide to buy the item at the announced price or the buyer can decide not to buy the item.

The main motivation for this work comes from the domain of online advertising. A large fraction of online advertisements (ads) are sold on online ad exchanges via auctions and this has led to a lot of research on revenue management in online advertising, see e.g. [25, 5, 6, 19, 2, 1, 23, 21]. Most ads are sold via second-price auctions, where the winner pays the second highest bid or a reserve price (whichever is larger), and no sale occurs if all of the bids are lower than the...
reserve price. However, a significant fraction of auctions only involve a single bidder [5, 6, 19] and this reduces to a posted-price auction when reserve prices announced: the seller sets a reserve price and the buyer decides whether to accept or reject it.

There are three main differences between this paper and previous work. First, unlike in previous work, we study repeated posted-price auctions from the perspective of the buyer that aims to maximize his expected utility, instead of from the perspective of the seller that aims to maximize his revenue. Second, previous papers assume that the buyer knows his valuation in each round. This valuation can either be a fixed value or an independently and identically distributed (i.i.d.) draw from a fixed distribution. In this paper, we do not make this assumption and study a version of the problem where the buyer does not know the distribution of his valuation and the buyer only observes a stochastic measurement of the valuation after he buys the item. Third, previous papers on dynamic pricing and auctions do not model imprecision and vagueness associated with the valuation of the item. In our setting, the buyer uses fuzzy sets to describe whether the valuation of the item is large or small. Furthermore, we use fuzzy sets to describe the dissatisfaction associated with the observed price. In our problem, the buyer makes decisions based on the probability of a fuzzy event [28]. His decision to buy or not depends on whether the satisfaction from having a high enough valuation for the item out-weights the dissatisfaction of the quoted price.

The goal of the buyer is to design a policy that has low regret, defined as the gap between the utility of a clairvoyant who has full information about all of the stochastic distributions and the utility achieved by a buyer facing an unknown distribution.

Our proposed algorithm uses Thompson Sampling [24] to balance the exploration-exploitation trade-off. Thompson Sampling is an idea that dates back to the work in [24] and has recently been analyzed theoretically in the context of multi-armed bandit problems [3, 4, 15]. To the best of our knowledge this technique has not been applied to exploration-exploitation trade-offs involving fuzzy sets and fuzzy events.

We summarize the main contributions of this paper as follows:

- To the best of our knowledge, we are the first to study exploration-exploitation trade-offs involving fuzzy sets and fuzzy events in the context of a dynamic pricing problem.
- We show how Thompson Sampling can be used to design tractable algorithms that can be used to dynamically learn the probability of fuzzy events over time.
- Experimental results show that our proposed method performs very well as the regret grows very slowly.

The remainder of this paper is organized as follows. In Section 2 we discuss the related literature. Section 3 provides a formal formulation of the problem. In Section 4 we present the our proposed algorithm. In Section 5 we perform experiments in order to assess the quality of our proposed algorithm. Section 6 concludes our work and provides some interesting directions for further research.
2 Related Literature

The problem considered in this paper is often referred to as a dynamic pricing problem in the operations research and management science community [9], and as a posted-price auction problem in the computer science community [17]. In the standard dynamic pricing problem there is a seller who wants to maximize revenue over some selling horizon by choosing prices in an optimal way. However, the precise relationship between price demand in unknown. This gives rise to the so-called exploration-exploitation trade-off. We refer the reader to [9] and [17] for a detailed overview of the dynamic pricing/posted-price auction problem. The main differences between this paper and existing works is as follows. First, most of the literature of dynamic pricing and learning focuses on perspective of the seller (see e.g. [17, 8, 16, 13, 22]). Similarly, most of the literature on posted-price auctions focuses on perspective of the seller (see e.g. [5, 6, 19]). However, in this paper, we focus on the buyer perspective. Second, most of the literature models the problem in a probabilistic way and does not consider combining fuzzy sets and online learning. While there are some papers that consider dynamic pricing and fuzzy logic such as [26, 12] and fuzzy demand [20, 14], these papers do not study exploration-exploitation trade-offs like we do in this paper.

From a methodological point of view, this paper is related to the literature on exploration-exploitation trade-offs, in particular, the literature related to multi-armed bandit problems [18, 10, 7]. In the traditional multi-armed bandit problem [18, 10] there is a finite set of actions, called arms, and each arm yields a stochastic reward. Play proceeds for a number of rounds, and in each round, precisely one arm can be selected. The goal in a multi-armed bandit problem is to learn which sequence of arms to select in order to maximize the expected cumulative reward over a number of rounds. Two of the main design principles for solving multi-armed bandit problems are (i) optimism in the face of uncertainty [7] and (ii) Thompson Sampling [24] or probability matching. Thompson Sampling is an idea that dates back to the work in [24]. Thompson Sampling has recently been analyzed theoretically in the context of multi-armed bandit problems [3, 4, 15]. Furthermore, this idea has been fruitfully applied in other online decision making problems, see e.g. [11]. However, to the best of our knowledge this technique has not been applied to exploration-exploitation trade-offs involving fuzzy sets and fuzzy events.

3 Problem Formulation

We consider a single buyer and a single seller that interact for $T$ rounds. An item is repeatedly offered for sale by the seller to the buyer over these $T$ rounds. In each round $t = 1, \ldots, T$, a price $p_t$ is offered by the seller and a decision $a_t \in \{0, 1\}$ is made by the buyer: $a_t$ takes value 1 when the buyer accepts to buy at that price, and $a_t$ takes value 0 otherwise. In
every round \( t \) there is a stochastic measurement \( v_t \in [0, 1] \) for the item. The value of \( v_t \) is an i.i.d. draw from a distribution \( D \) and has expectation \( \nu = \mathbb{E}\{v_t\} \). The measurement \( v_t \) is only revealed to the buyer if he buys the item in round \( t \), i.e., the buyer only observes the value of the measurement after he buys the item. We assume that the buyer does not know the distribution \( D \).

From the buyer perspective we assume that there is imprecision and vagueness associated with the valuation of the item. We assume that the buyer uses \( M_v \in \mathbb{N} \) fuzzy sets [27, 29] to describe his valuation for the item. Denote these fuzzy sets by \( V^1, \ldots, V^{M_v} \). The membership function is given by \( \mu^m_v(x) \) for \( m = 1, \ldots, M_v \) and maps a stochastic measurement to values in \([0, 1]\). For example, the buyer could use three fuzzy sets that describe valuations that can be low, medium or high.

We assume that the buyer also uses fuzzy sets to describe his dissatisfaction with a particular price. More specifically, the buyer uses \( M_p \in \mathbb{N} \) fuzzy sets to describe his dissatisfaction for buying the item at a particular price. Denote these fuzzy sets by \( D^1, \ldots, D^{M_p} \). The membership function is given by \( \mu^k_p(x) \) for \( k = 1, \ldots, M_p \) and maps the price to values in \([0, 1]\). For example, the buyer could use three fuzzy sets to describe that his dissatisfaction can be low, medium or high.

We make the following assumption on the membership functions.

**Assumption 1** The membership functions \( \mu^m_v(x) \) for \( m = 1, \ldots, M_v \) and \( \mu^k_p(x) \) for \( k = 1, \ldots, M_p \) are fixed for all rounds \( t = 1, \ldots, T \).

In our problem, the buyer makes decisions based on the probability of a fuzzy event. The definition for the probability of a fuzzy event is given by Definition 1 and is due to [28].

**Definition 1 ([28]).** Let \( A \) be an arbitrary fuzzy set with membership function \( \mu_A(x) \). The probability of the fuzzy event associated with the fuzzy set \( A \) with respect to a distribution \( D \) is defined as \( \mathbb{P}\{A\} = \mathbb{E}_D \{\mu_A\} \). Here \( \mathbb{E}_D \) denotes the expectation under distribution \( D \).

If the buyer had complete information about the distribution \( D \), then he can calculate the probability of the fuzzy events \( \mathbb{P}\{V^m\} \) for \( m = 1, \ldots, M_v \) associated with the fuzzy sets \( V^1, \ldots, V^{M_v} \). We assume that the buyer uses a function \( F^V : [0, 1]^{M_v} \to [0, 1] \) that combines the probabilities \( \mathbb{P}\{V^m\} \) for \( m = 1, \ldots, M_v \) into an aggregated score \( S_V \in [0, 1] \).

Similarly, we assume that the buyer uses a function \( F^P : [0, 1]^{M_p} \to [0, 1] \) that combines membership values for the fuzzy sets \( D^1, \ldots, D^{M_p} \) into an aggregated score \( S_P \in [0, 1] \).

If the buyer had complete information about the distribution \( D \), we assume that the buyer makes decisions according to the rule described in Assumption 2.

**Assumption 2** Under complete information about the distribution \( D \), the buyer uses the following rule to make decisions:
Dynamic pricing using Thompson Sampling with fuzzy events

- observe the price $p_t$ in round $t$.
- calculate $S_P(p_t) = F^P(\mu_1^P(p_t), \ldots, \mu_M^P(p_t))$.
- compare the value of $S_P(p_t)$ with the value of $S_V = F^V(\mathbb{P}\{V^1\}, \ldots, \mathbb{P}\{V^M\})$.
- if $S_V \geq S_P(p_t)$, then buy the item.
- if $S_V < S_P(p_t)$, then do not buy the item.

The intuition behind this rule is that the value of $S_V$ represents the total aggregated degree of satisfaction with the item and that $S_P(p_t)$ represents the total aggregated dissatisfaction with the price $p_t$. The decision rule described above indicates that the buyer would only buy the item if his total satisfaction with the item outweighs the dissatisfaction at price $p_t$.

Note that the value of $S_V$ depends on the function $F^V$ and the probabilities $\mathbb{P}\{V^1\}, \ldots, \mathbb{P}\{V^M\}$: if the same function $F^V$ is used in all the rounds, then $S_V$ is fixed. Furthermore, note that the value of $S_V$ is unknown to the buyer since the distribution $D$ is unknown.

The expected utility of the buyer in round $t$ is given by $u_t = a_t \cdot (S_V - S_P(p_t))$. In other words, if the buyer purchases the item ($a_t = 1$) the utility is the difference between the total aggregated degree of satisfaction and the total aggregated dissatisfaction with the price. Otherwise, the utility is zero. The objective of the buyer is to maximize his expected utility over the $T$ rounds.

For a fixed sequence $p_1, \ldots, p_T$ of observed prices and a fixed sequence of decisions $a_1, \ldots, a_T$ by the buyer, the regret of the buyer over $T$ rounds is defined as

$$R_T = \sum_{t=1}^{T} \max\{S_V - S_P(p_t), 0\} - \sum_{t=1}^{T} a_t \cdot (S_V - S_P(p_t)).$$

The expected regret over $T$ rounds is defined as

$$\mathbb{R}_T = \mathbb{E}\{R_T\},$$

where the expectation in Equation (2) is taken with respect to possible randomization in the selection of the actions $a_1, \ldots, a_T$.

Note that (for a fixed sequence $p_1, \ldots, p_T$ of observed prices) the objective of maximizing expected utility is equivalent to minimizing the expected regret. The goal is to make decisions that are as close as possible to the decisions that are prescribed by the decision rule in Assumption 2. Since the distribution $D$ is unknown to the buyer, he faces an exploration-exploitation trade-off. In order to gain information and estimate the value of $S_V$ he needs to buy the item a number of times (exploration). However, he also wants to use all information collected so far and only buy the item if $S_V \geq S_P(p_t)$ holds (exploitation). In this paper, we seek to develop algorithms that ensure that the value of $\mathbb{R}_T$ grows slowly as the problem horizon $T$ increases.
4 Proposed algorithm

In this section we discuss our proposed algorithm. We refer to our algorithm as TS-PFE (Thompson Sampling for Probabilities of Fuzzy Events). Our algorithm uses Thompson Sampling [24] to balance the exploration-exploitation trade-off. The pseudo-code for TS-PFE is given by Algorithm 1.

Algorithm 1: TS-PFE

Require: Parameters of prior distribution \((a^m_0, b^m_0)\) for \(m = 1, \ldots, M_v\), function \(F^V\), function \(F^P\), membership functions \(\mu^m_v(x)\) for \(m = 1, \ldots, M_v\) and \(\mu^k_p(x)\) for \(k = 1, \ldots, M_p\).

1: for \(m = 1 \text{ to } M_v\) do
2: \hspace{1em} Set \(a^m = a^m_0\). Set \(b^m = b^m_0\).
3: end for
4: Set \(t = 1\).
5: Purchase item in first round.
6: Observe \(v_1\).
7: for \(m = 1 \text{ to } M_v\) do
8: \hspace{1em} Set \(a^m = a^m + \mu^m_v(v_1)\). Set \(b^m = b^m + (1 - \mu^m(v_1))\).
9: end for
10: for \(t \in \{2, \ldots, T\}\) do
11: \hspace{1em} Observe price \(p_t\) of item.
12: \hspace{1em} for \(m = 1 \text{ to } M_v\) do
13: \hspace{2em} Sample \(\theta^m \sim B(a^m, b^m)\).
14: \hspace{1em} end for
15: \hspace{1em} Set \(S^V = F^V(\hat{\theta}^1, \ldots, \hat{\theta}^{M_v})\).
16: \hspace{1em} Set \(S^P(p_t) = F^P(\mu^1_p(p_t), \ldots, \mu^{M_p}_p(p_t))\).
17: \hspace{1em} if \(S^V \geq S^P(p_t)\) then
18: \hspace{2em} Purchase item at price \(p_t\).
19: \hspace{1em} Observe \(v_t\).
20: \hspace{1em} for \(m = 1 \text{ to } M_v\) do
21: \hspace{2em} Set \(a^m = a^m + \mu^m_v(v_t)\). Set \(b^m = b^m + (1 - \mu^m(v_t))\).
22: \hspace{1em} end for
23: \hspace{1em} end if
24: end for

Our algorithm adopts a Bayesian framework for handling the unknown probabilities of fuzzy events \(P\{V^1\}, \ldots, P\{V^{M_v}\}\). Let \(Z \sim B(a, b)\) denote a random variable \(Z\) that follows a Beta distribution with parameters \(a\) and \(b\) and with expectation \(\mathbb{E}\{Z\} = \frac{a}{a+b}\).

We associate a Beta distributed random variable \(\theta^m\) with the probability \(P\{V^m\}\) for \(m = 1, \ldots, M_v\). At the start of our algorithm, \(\theta^m \sim B(a^m_0, b^m_0)\) indicates the prior distribution. Every time that the buyer buys the item, he
observes an i.i.d. draw $v_t$ from the distribution $D$. This draw is subsequently used to update the distribution of $\theta^m$. We can use the draw $v_t$ in order to learn the membership values corresponding to the fuzzy sets at the value of $v_t$. We subsequently use this information to update our estimate of $P\{V^m\}$ for $m = 1, \ldots, M_v$.

In order to decide to buy the item or not, we sample values $\hat{\theta}^m$ from the posterior distributions $B(a^m, b^m)$ and use these sampled values in order to form an estimate $\hat{S}_V$ of $S_V$. This is the Thompson Sampling step. Intuitively, the buyer buys the item with probability $P_t\{S_V \geq S_P(p_t)\}$ and where the probability measure $P_t$ is induced by the posterior distribution $B(a^m, b^m)$ in round $t$. It is precisely this Thompson Sampling step that balances exploration and exploitation.

5 Experiments

In this section we conduct experiments in order to test the performance of our proposed algorithm. In total we have three experimental settings, which we describe in more detail below.

5.1 General settings and performance metrics

In all our experiments the prior parameters are set to $(a^m_0 = 1, b^m_0 = 1)$ for $m = 1, \ldots, M_v$ which corresponds to a uniform distribution on $[0, 1]$. The problem horizon is set to $T = 5000$.

In order to measure the performance of the methods, we consider two performance metrics. Our main performance metric is the cumulative regret which is defined as $R_T = \sum_{t=1}^{T} \max\{S_V - S_P(p_t), 0\} - \sum_{t=1}^{T} a_t \cdot (S_V - S_P(p_t))$.

The second performance metric is the fraction of rounds in which the “best action” was selected. Here “best action” means the action that the buyer would have taken if he had complete information about the distribution $D$. That is, if the buyer makes decisions according to the rule described in Assumption 2.

We run 500 independent simulations and all performance metrics are averaged over these 500 simulations.

5.2 Experimental Setting I

In this setting $v_t$ is drawn from an uniform distribution on $[a - 0.3, a + 0.3]$, where $a = 0.5$. The price $p_t$ is an i.i.d. draw from the distribution specified in Table 1 with $a = 0.5$.

We use three fuzzy sets $V^1, V^2, V^3$ for the valuation of the buyer and so $M_v = 3$. The three fuzzy sets describe valuations that can be low ($V^1$), medium ($V^2$) or high ($V^3$). We use a single fuzzy set to express the dissatisfaction with the price and so $M_p = 1$. The membership functions are displayed in Figure 1 and Figure 2. The interpretation is that for high prices the membership value will
be 1 and the buyer is very dissatisfied. For intermediate values the membership value is between 0 and 1, and indicates that the buyer is partially dissatisfied. For low values of the price, the buyer is not dissatisfied.

For the function $F^V$ we use $F^V = 0.25 \cdot \mathbb{P}\{V^2\} + 0.75 \cdot \mathbb{P}\{V^3\}$. For the function $F^P$ we use $F^P = \mu^1_p(p_t)$.

| Value   | Probability |
|---------|-------------|
| $a - 0.2$ | $\frac{1}{11}$ |
| $a - 0.1$ | $\frac{1}{11}$ |
| $a - 0.05$ | $\frac{1}{11}$ |
| $a - 0.02$ | $\frac{1}{11}$ |
| $a$ | $\frac{1}{11}$ |
| $a + 0.05$ | $\frac{1}{11}$ |
| $a - 0.2$ | $\frac{1}{11}$ |
| $a - 0.3$ | $\frac{1}{11}$ |

Table 1. Distribution of prices.

The results for the cumulative regret and the fraction of the rounds in which the best action is selected are displayed in Figure 3 and Figure 4. The results indicate that TS-PFE performs relatively well as the cumulative regret grows very slowly as the number of rounds increases. In Figure 4 we can see that TS-PFE learns relatively quickly as it selects the "best action" in at least 90% of the first 500 rounds. Furthermore, this percentage only increases as the number of rounds increases indicating that TS-PFE makes less and less mistakes.
5.3 Experimental Setting II

In this setting $v_t$ is drawn from an uniform distribution on $[a - 0.3, a + 0.3]$, where $a = 0.3$. The price $p_t$ is an i.i.d. draw from the distribution specified in Table 1 with $a = 0.3$. The membership function used to express the dissatisfaction with the price is different compared to experimental setting I and is displayed in Figure 5. The membership functions for $V^1, V^2, V^3$ are the same as in experimental setting I and are displayed in Figure 1. All other settings are the same as in experimental setting I.

The results for the cumulative regret and the fraction of the rounds in which the best action is selected are displayed in Figure 6 and Figure 7. The results are qualitatively similar as those reported in Figure 3 and Figure 4. Again we see that TS-PFE is able to learn relatively quickly to make the correct decisions.
5.4 Experimental Setting III

In this setting we test how the algorithm performs when the function $F_V$ changes during the problem horizon. In this setting $v_t$ is drawn from an uniform distribution on $[a - 0.3, a + 0.3]$, where $a = 0.3$. The price $p_t$ is an i.i.d. draw from the distribution specified in Table 2 with $a = 0.3$. The membership function used to express the dissatisfaction with the price is different compared to experimental setting I and is displayed in Figure 5. The membership functions for $V^1, V^2, V^3$ are the same as in experimental setting I and are displayed in Figure 1.

For the function $F_V$ we use $F_V = 0.25 \cdot \mathbb{P}\{V^2\} + 0.75 \cdot \mathbb{P}\{V^3\}$ in rounds $t < 2500$ and we take $F_V$ to be the harmonic mean of $\mathbb{P}\{V^2\}$ and $\mathbb{P}\{V^3\}$ in rounds $t \geq 2500$. The problem horizon is set to $T = 10000$. All other settings are the same as in experimental setting I. With these settings the optimal action is not to buy in rounds $t < 2500$ and to buy at price 0.30 in rounds $t \geq 2500$.

The results for the cumulative regret and the fraction of the rounds in which the best action is selected are displayed in Figure 8 and Figure 9. The results are qualitatively similar as those reported in Figure 3 and Figure 4. The level of regret is in general higher in Figure 8 compared to Figure 3 and 6, which indicates the the problem is harder. Also, the percentiles in Figures 8 and 9 indicate that the problem is harder. However, the overall pattern of regret is similar to the other experimental settings. Again we see that TS-PFE is able to learn relatively quickly to make the correct decisions, even if the function $F_V$ changes during the problem horizon.
Table 2. Distribution of prices.

| Value | Probability |
|-------|-------------|
| 0.3   |             |
| 0.5   | 1           |

Fig. 8. Cumulative regret for experimental setting III averaged over 500 runs. Solid lines indicate the mean and shaded region indicates 5-th and 95-th percentiles.

Fig. 9. Fraction of rounds in which the optimal action is selected for experimental setting III averaged over 500 runs. Solid lines indicate the mean and shaded region indicates 5-th and 95-th percentiles.

6 Conclusion

In this paper we study a repeated posted-price auction between a single seller and a single buyer that interact for a finite number of periods or rounds. In this paper we study the problem from the perspective of the buyer who only gets to observe a stochastic measurement of the valuation of the item after he buys the item. In our model, the buyer uses fuzzy sets to describe his satisfaction with the observed valuations and he uses fuzzy sets to describe his dissatisfaction with the observed price. In our problem, the buyer makes decisions based on the probability of a fuzzy event. His decision to buy or not depends on whether the satisfaction from having a high enough valuation for the item out-weights the dissatisfaction of the quoted price. To the best of our knowledge, we are the first to study exploration-exploitation trade-offs involving fuzzy sets and fuzzy events in the context of a dynamic pricing problem. We show how Thompson Sampling can be used to design tractable algorithms that can be used to dynamically learn the probability of fuzzy events over time.

One direction for future work is to investigate whether the ideas used in this paper can be used in other exploration-exploitation problems involving fuzzy sets and probabilities based on fuzzy events. For example, it would be interesting to investigate whether the ideas can be extended or adapted to a setting of group decision making with stochastic feedback and preferences described by fuzzy sets. One could for example try to learn to make decisions online such that the preferences of the members of the group are/ remain close to each other.
Acknowledgements. We would like to thank Rick Augustinus Maria Gilsing, Jonro Erasmus, Anasztázia Junger, Peipei Chen, Bambang Suratno, and Onat Ege Adalı for interesting discussions, which helped improve this paper considerably.

References

1. Afshar, R.R., Zhang, Y., Firat, M., Kaynak, U.: A decision support method to increase the revenue of ad publishers in waterfall strategy. In: 2019 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr). pp. 1–8 (May 2019). https://doi.org/10.1109/CIFEr.2019.8759106
2. Afshar., R.R., Zhang., Y., Firat., M., Kaymak., U.: A reinforcement learning method to select ad networks in waterfall strategy. In: Proceedings of the 11th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,. pp. 256–265. INSTICC, SciTePress (2019). https://doi.org/10.5220/0007395502560265
3. Agrawal, S., Goyal, N.: Analysis of thompson sampling for the multi-armed bandit problem. In: Mannor, S., Srebro, N., Williamson, R.C. (eds.) Proceedings of the 25th Annual Conference on Learning Theory. Proceedings of Machine Learning Research, vol. 23, pp. 39.1–39.26. PMLR, Edinburgh, Scotland (25–27 Jun 2012), http://proceedings.mlr.press/v23/agrawal12.html
4. Agrawal, S., Goyal, N.: Further optimal regret bounds for thompson sampling. In: Carvalho, C.M., Ravikumar, P. (eds.) Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 31, pp. 99–107. PMLR, Scottsdale, Arizona, USA (29 Apr–01 May 2013), http://proceedings.mlr.press/v31/agrawal13a.html
5. Amin, K., Rostamizadeh, A., Syed, U.: Learning prices for repeated auctions with strategic buyers. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. pp. 1169–1177. NIPS’13, Curran Associates Inc., USA (2013), http://dl.acm.org/citation.cfm?id=2999611.2999742
6. Amin, K., Rostamizadeh, A., Syed, U.: Repeated contextual auctions with strategic buyers. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 622–630. Curran Associates, Inc. (2014), http://papers.nips.cc/paper/5589-repeated-contextual-auctions-with-strategic-buyers.pdf
7. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multi-armed bandit problem. Machine Learning 47(2), 235–256 (May 2002). https://doi.org/10.1023/A:1013689704352, https://doi.org/10.1023/A:1013689704352
8. Besbes, O., Zeevi, A.: On the (surprising) sufficiency of linear models for dynamic pricing with demand learning. Management Science 61(4), 723–739 (2015). https://doi.org/10.1287/mnsc.2014.2031, https://doi.org/10.1287/mnsc.2014.2031
9. den Boer, A.V.: Dynamic pricing and learning: Historical origins, current research, and new directions. Surveys in Operations Research and Management Science 20(1), 1 – 18 (2015). https://doi.org/https://doi.org/10.1016/j.sorms.2015.03.001, http://www.sciencedirect.com/science/article/pii/S1876735415000021
10. Bubeck, S., Cesa-Bianchi, N.: Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Foundations and Trends® in Machine Learning 5(1), 1–122 (2012). https://doi.org/10.1561/2200000024, http://dx.doi.org/10.1561/2200000024
11. Chapelle, O., Li, L.: An empirical evaluation of Thompson sampling. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. pp. 2249–2257. NIPS’11, Curran Associates Inc., USA (2011), http://dl.acm.org/citation.cfm?id=2986459.2986710

12. Gen-dao, L., Wei, L.: Dynamic pricing of perishable products with random fuzzy demand. In: 2010 International Conference on Management Science Engineering 17th Annual Conference Proceedings. pp. 191–199 (Nov 2010). https://doi.org/10.1109/ICMSE.2010.5719804

13. Harrison, J.M., Keskin, N.B., Zeevi, A.: Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution. Management Science 58(3), 570–586 (2012). https://doi.org/10.1287/mnsc.1110.1426, https://doi.org/10.1287/mnsc.1110.1426

14. Kao, C., Hsu, W.K.: A single-period inventory model with fuzzy demand. Computers & Mathematics with Applications 43(6), 841 – 848 (2002). https://doi.org/https://doi.org/10.1016/S0898-1221(01)00325-X, http://www.sciencedirect.com/science/article/pii/S089812210100325X

15. Kaufmann, E., Korda, N., Munos, R.: Thompson sampling: An asymptotically optimal finite-time analysis. In: Bshouty, N.H., Stoltz, G., Vayatis, N., Zeugmann, T. (eds.) Algorithmic Learning Theory. pp. 199–213. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)

16. Keskin, N.B., Zeevi, A.: Chasing demand: Learning and earning in a changing environment. Mathematics of Operations Research 42(2), 277–307 (2017). https://doi.org/10.1287/moor.2016.0807, https://doi.org/10.1287/moor.2016.0807

17. Kleinberg, R., Leighton, T.: The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science. pp. 594–. FOCS ’03, IEEE Computer Society, Washington, DC, USA (2003), http://dl.acm.org/citation.cfm?id=946243.946352

18. Lai, T.L., Robbins, H.: Asymptotically efficient adaptive allocation rules. Advances in applied mathematics 6(1), 4–22 (1985)

19. Mohri, M., Medina, A.M.n.: Optimal regret minimization in posted-price auctions with strategic buyers. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, pp. 1871–1879. NIPS'14, MIT Press, Cambridge, MA, USA (2014), http://dl.acm.org/citation.cfm?id=2969033.2969036

20. Petrović, D., Petrović, R., Vujšević, M.: Fuzzy models for the newsboy problem. International Journal of Production Economics 45(1), 435 – 441 (1996). https://doi.org/https://doi.org/10.1016/0925-5273(96)00014-X, http://www.sciencedirect.com/science/article/pii/092552739600014X, proceedings of the Eighth International Symposium on Inventories

21. Rhuggenaath, J., Akcay, A., Zhang, Y., Kaymak, U.: Optimizing reserve prices for publishers in online ad auctions. In: 2019 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr). pp. 1–8 (May 2019). https://doi.org/10.1109/CIFEr.2019.8759123

22. Rhuggenaath, J., de Oliveira da Costa, P.R., Akcay, A., Zhang, Y., Kaymak, U.: A heuristic policy for dynamic pricing and demand learning with limited price changes and censored demand. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). pp. 3693–3698 (Oct 2019). https://doi.org/10.1109/SMC.2019.8914590
23. Rhuggenaath, J., Akcay, A., Zhang, Y., Kaymak, U.: Optimal display-ad allocation with guaranteed contracts and supply side platforms. Computers & Industrial Engineering 137, 106071 (2019). https://doi.org/https://doi.org/10.1016/j.cie.2019.106071, http://www.sciencedirect.com/science/article/pii/S0360835219305303

24. THOMPSON, W.R.: On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika 25(3-4), 285–294 (1933). https://doi.org/10.1093/biomet/25.3-4.285, http://dx.doi.org/10.1093/biomet/25.3-4.285

25. Wang, J., Zhang, W., Yuan, S.: Display advertising with real-time bidding (RTB) and behavioural targeting. Foundations and Trends® in Information Retrieval 11(4-5), 297–435 (2017). https://doi.org/10.1561/1500000049, http://dx.doi.org/10.1561/1500000049

26. Xiong, Y., Li, G., Fernandes, K.J.: Dynamic pricing model and algorithm for perishable products with fuzzy demand. Applied Stochastic Models in Business and Industry 26(6), 758–774 (2010). https://doi.org/10.1002/asmb.816, https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.816

27. Zadeh, L.: Fuzzy sets. Information and Control 8(3), 338 – 353 (1965). https://doi.org/https://doi.org/10.1016/S0019-9958(65)90241-X, http://www.sciencedirect.com/science/article/pii/S001999586590241X

28. Zadeh, L.: Probability measures of fuzzy events. Journal of Mathematical Analysis and Applications 23(2), 421 – 427 (1968). https://doi.org/https://doi.org/10.1016/0022-247X(68)90078-4, http://www.sciencedirect.com/science/article/pii/0022247X68900784

29. Zadeh, L.: The concept of a linguistic variable and its application to approximate reasoning—i. Information Sciences 8(3), 199 – 249 (1975). https://doi.org/https://doi.org/10.1016/0020-0255(75)90039-5, http://www.sciencedirect.com/science/article/pii/0020025575900385