Dynamic Contextual Pricing with Doubly Non-Parametric Random Utility Models

Elynn Chen[1], Lan Gao[2], Jiayu Li[1]

[1]New York University, [2]University of Tennessee Knoxville

Objectives

The paper explores the use of doubly non-parametric random utility models in dynamic contextual pricing, where both the mean utility and the market noise are modeled in a non-parametric manner.

- Propose Innovative Estimation Techniques: The estimation of these random utility models is particularly complex as the quantities of interest, including consumers' true utility and the underlying market noise, are not directly observable. To address these challenges, we employ Distributional Nearest Neighbour (DNN) and two-scaled Distributional Nearest Neighbour (TDNN) estimation for the mean utility function and a kernel estimation for the distribution of the market noise, offering a more robust framework for understanding consumer demand in dynamic pricing scenarios.
- Theoretical Contributions in Statistical Learning: We establish the regret bound for our doubly non-parametric procedure for dynamic pricing. The regret analysis relied on a finite-sample uniform convergence rate for the DNN and TDNN estimators ([1]) of $\mathring{\mu}(x)$, which is new to the literature and can be of independent interest. Under mild conditions, we show that for a market noise c.d.f. $F(\cdot)$ with m-th order derivative ($m \geqslant 3$), our policy achieves a regret upper bound of $\tilde{\mathcal{O}}_d(T^{\frac{d+4}{d+8}})$, where T is time horizon and $\tilde{\mathcal{O}}_d$ is the order that hides logarithmic terms and the dimensionality of feature is $d(d \geqslant 3)$.

Introduction

We consider a contexual dynamic pricing problem where a seller sells a single product over a horizon of T periods. At each period $t \in \{1, 2, \ldots, T\}$, the seller observes a d-dimensional contextual covariate $\mathbf{x}_t \in \mathbb{R}^d$, which encodes the product and customer information, and other information that can be revealed to the seller. Then the expected market value of the product at period t is determined by the observed contextual covariate $\mathbf{x}_t$ through a general random utility model:

$$v_t = \mathring{\mu}(\mathbf{x}_t) + \epsilon_t, \tag{1}$$

where $\mathring{\mu}(\cdot)$ is the *unknown function* of the mean utility, and ϵ_t are i.i.d. noises following an *unknown distribution* $F(\cdot)$ with $\mathbb{E}[\epsilon_t] = 0$. Given a posted price of p_t for the product at period t, we observe $y_t := \mathbf{1}(v_t \geqslant p_t)$ that indicates whether a sale occurs ($y_1 = 1$) or not ($y_1 = 0$). Specifically, the model is equivalent to the probabilistic model:

$$y_t = \begin{cases} 0, & \text{with probability} \quad F(p_t - \mathring{\mu}(\mathbf{x}_t)), \\ 1, & \text{with probability} \quad 1 - F(p_t - \mathring{\mu}(\mathbf{x}_t)). \end{cases} \tag{2}$$

Therefore, given a posted price p_t and the contextual covariate $\mathbf{x}_t$, the expected revenue at period t is defined as

$$\mathrm{rev}_t(p_t) := \mathbb{E}[p_t \mathbf{1}(y_t = 1)] = p_t \cdot (1 - F(p_t - \mathring{\mu}(\mathbf{x}_t))).$$

Meanwhile, if the mean utility function $\mathring{\mu}(\cdot)$ and the cumulative distribution $F(\cdot)$ are known, the optimal price should be set as

$$p_t^* := \arg\max_p \left[p\left(1 - F(p - \mathring{\mu}(\mathbf{x}_t))\right)\right].$$

For a policy π that sets price p_t at period t for $t \geqslant 1$, its regret over the horizon of T periods is defined as

$$\mathrm{Reg}(T; \pi) := \sum_{t=1}^{T} \mathrm{rev}_t(p_t^*) - \mathrm{rev}_t(p_t). \tag{3}$$

The goal of a decision maker is to design a pricing policy π^* that minimizes $\mathrm{Reg}(T; \pi)$, or equivalently, maximize the accumulative expected revenue $\sum_{t=1}^{T} \mathrm{rev}_t(p_t)$.

What sets this paper apart from existing literature is that we *do not* impose any parametric form on either the mean utility $\mathring{\mu}(\cdot)$ or the unknown distribution $F(\cdot)$, thus eliminating the risk of model misspecification. Here we aim to estimate them both non-parametrically while also deriving an optimal pricing policy with theoretical guarantees. This task presents unique challenges because everything, i.e. $\mathring{\mu}(\cdot)$, or ϵ_t, in model (1) are not directly observable.

Method

We propose Algorithm 1 which outputs the empirical optimal policy for contextual dynamic pricing and strikes a balance of the exploration-exploitation trade-off. The horizon of T periods is divided into episodes with increasing length. Specifically, the kth episode has length $n_k = 2^{k-1} n_0$, where $k \geqslant 1$ and n_0 is the minimal episode length. For each $k \geqslant 1$, the kth episode is further divided into exploration and exploitation phases, which are denoted by $\mathcal{T}_{k,exp}$ and $\mathcal{T}_{k,com}$, respectively.

Mean Utility Function Estimation: To address the challenge of not directly observing random utility $\mathring{\mu}(\cdot)$, we observe the following property for single-product dynamic pricing. Specifically, if we adopt the uniform random policy $p_t \sim \mathbf{U}(0, B)$, we can recover the mean utility function in an average sense. Denote by $g_t := B y_t$. It can be observed that:

$$\mathbb{E}[g_t \,|\, \mathbf{x}_t] = \mathbb{E}_{\epsilon_t}\Big[\mathbb{E}_{p_t}[B y_t \,|\, \mathbf{x}_t, \epsilon_t] \mathbf{x}_t\Big] = B\mathbb{E}_{\epsilon_t}\Big[\frac{\mathring{\mu}(\mathbf{x}_t) + \epsilon_t}{B} \Big| \mathbf{x}_t\Big] = \mathring{\mu}(\mathbf{x}_t). \tag{4}$$

Therefore, if the seller randomly sets i.i.d. price $p_t \sim \mathbf{U}(0, B)$ for each period t in the exploration phase $\mathcal{T}_{k,exp}$, then the mean utility function $\mathring{\mu}(\mathbf{x}_t)$ can be estimated by regressing g_t on $\mathbf{x}_t$ based on observations $\mathbf{z}_t := \{(\mathbf{x}_t, g_t)\}_{t \in \mathcal{T}_{k,exp}}$. The **DNN** $\hat{\mu}_k(\mathbf{x}; s)$ with sub-sampling scale s for estimating $\mathring{\mu}(\mathbf{x})$ is formally defined as the following U-statistic

$$\hat{\mu}_k(\mathbf{x}; s) = \binom{n_{k,exp}}{s}^{-1} \sum_{1 \leqslant i_1 \leqslant \cdots \leqslant i_s \leqslant n_{k,exp}} g_{(1)}(\mathbf{z}_{i_1}, \cdots, \mathbf{z}_{i_s}), \tag{5}$$

where $g_{(1)}(\mathbf{z}_{i_1}, \cdots, \mathbf{z}_{i_s})$ is the 1-NN estimator for the subsample $(\mathbf{z}_{i_1}, \cdots, \mathbf{z}_{i_s})$. Formally, it returns the value of $g_{(1)}(\mathbf{z}_{i_1}, \cdots, \mathbf{z}_{i_s}) = g_{(1)}$ from the reordered sample $\{(\mathbf{x}_{(1)}, g_{(1)}), \cdots, (\mathbf{x}_{(s)}, g_{(s)})\}$ of $\{(\mathbf{x}_{i_1}, g_{i_1}), \ldots, (\mathbf{x}_{i_s}, g_{i_s})\}$ such that

$$\left\|\mathbf{x}_{(1)} - \mathbf{x}\right\| \leqslant \left\|\mathbf{x}_{(2)} - \mathbf{x}\right\| \leqslant \cdots \leqslant \left\|\mathbf{x}_{(s)} - \mathbf{x}\right\|, \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm.

The **TDNN**(two-scale DNN) can be applied to remove the first-order bias of the DNN estimator when the mean utility function and density function of covariates $\mathbf{x}_t$ have higher level of smoothness. Specifically, a TDNN estimator is given by

$$\hat{\mu}_k(\mathbf{x}; s_1, s_2) = \alpha_1 \hat{\mu}_k(\mathbf{x}; s_1) + \alpha_2 \hat{\mu}_k(\mathbf{x}; s_2), \tag{7}$$

where α_1 and α_2 satisfy $\alpha_1 + \alpha_2 = 1$ and $\alpha_1 s_1^{-2/d} + \alpha_2 s_2^{-2/d} = 0$.

Noise Distribution Estimation: Based on the observation that

$$F(z) = \mathbb{P}(\epsilon_t \leqslant z) = 1 - \mathbb{E}[y_t | p_t - \mathring{\mu}(\mathbf{x}_t) = z], \tag{8}$$

$F(\cdot)$ can be estimated by the Nadaraya-Watson kernel regression estimator([2]). Given sample $\{\mathbf{x}_t, y_t, p_t\}_{t \in \mathcal{T}_{k,exp}}$ collected in the exploration phase and any estimated utility function $\mu(\cdot)$, we define

$$\begin{aligned} \hat{a}_k(z; \mu) &:= n_{k,exp} b_k^{-1} \sum_{t \in \mathcal{T}_{k,exp}} K\left(\frac{p_t - \mu(\mathbf{x}_t) - z}{b_k}\right) y_t, \\ \hat{\xi}_k(z; \mu) &:= n_{k,exp} b_k^{-1} \sum_{t \in \mathcal{T}_{k,exp}} K\left(\frac{p_t - \mu(\mathbf{x}_t) - z}{b_k}\right), \end{aligned} \tag{9}$$

where $K(\cdot)$ is a chosen m-th order kernel function and b_k is the bandwidth that will be chosen appropriately later. Then the Nadaraya-Watson kernel estimator of $F(z)$ is defined by

$$\hat{F}_k(z; \mu) = 1 - \frac{\hat{a}_k(z; \mu)}{\hat{\xi}_k(z; \mu)}. \tag{10}$$

Feasible Optimal Decision. In reality, we can only estimate $\mathring{\mu}(\cdot)$ and $F(\cdot)$ from collected data. In the k-th episode of Algorithm 1, we obtain the estimated mean utility function $\hat{\mu}_k$ using either the DNN or TDNN estimator and calculate the Nadaraya-Watson estimator $\hat{F}_k(z)$ based on (10). Accordingly, the optimal offer price $\hat{p}_t$ for $t \in \mathcal{T}_{k,com}$ is obtained by optimizing the estimated reward. That is, $\hat{p}_t = \arg\max_{p_t > 0} p_t[1 - \hat{F}_k(p_t - \hat{\mu}_k(\mathbf{x}_t))]$. As a result, we have that for any $t \in \mathcal{T}_{k,com}$,

$$\hat{p}_t = \hat{h}_k \circ \hat{\mu}_k(\mathbf{x}_t), \tag{11}$$

where

$$\hat{h}_k(v) = v + \hat{\phi}_k^{-1}(-v) \quad \text{and} \quad \hat{\phi}_k(z) = z - \frac{1 - \hat{F}_k(z)}{\hat{F}_k^{(1)}(z)}. \tag{12}$$

Here $\hat{F}_k^{(1)}(z)$ represents the first-order derivative of $\hat{F}_k(z)$ and is defined by

$$\hat{F}_k^{(1)}(z) = \hat{F}_k^{(1)}(z; \hat{\mu}_k) = -\frac{\hat{a}_k^{(1)}(z; \hat{\mu}_k)\hat{\xi}_k(z; \hat{\mu}_k) - \hat{a}_k(z; \hat{\mu}_k)\hat{\xi}_k^{(1)}(z; \hat{\mu}_k)}{\hat{\xi}_k(z; \hat{\mu}_k)^2}, \tag{13}$$

where $\hat{a}_k(z; \hat{\mu}_k)$ and $\hat{\xi}_k(z; \hat{\mu}_k)$ are defined in (9) and their corresponding first derivatives are easy to be calculated accordingly.

Algorithm

Algorithm 1: Non-Parametric Contextual Pricing

Input: Upper bound $B > 0$ for the market values, and minimal episode length n_0.

Calculate the length of the k-th episode $n_k = 2^{k-1} n_0$, and the length of k-th exploration episode $n_{k,exp}$.
Let $\mathcal{T}_{k,exp} := \{n_k - n_0 + 1, \cdots, n_k - n_0 + n_{k,exp}\}$ and $\mathcal{T}_{k,com} := \{n_k - n_0 + n_{k,exp} + 1, \cdots, n_{k+1} - n_0\}$.

`/* ******* Episodic Learning and Decision ******* */`

for *each episode* $k = 1, 2, \ldots$ **do**

`/* ******* Exploration on` $\mathcal{T}_{k,exp}$ `******* */`

for $t \in \mathcal{T}_{k,exp}$ **do**

Offer price $p_t \sim U(0, B)$ and collect data $\{(\mathbf{x}_t, By_t)\}_{t \in \mathcal{T}_{k,exp}}$;

`/* ******* Estimation ******* */`

Using samples $\{(\mathbf{x}_t, By_t)\}_{t \in \mathcal{T}_{k,exp}}$ to fit $\hat{\mu}_k(\mathbf{x}_t)$ for $t \in \mathcal{T}_{k,exp}$ according to DNN or TDNN (7).

Update estimates of the error distribution $\hat{F}_k(u)$ by the Nadaraya-Watson kernel estimator (10).

`/* ******* Commitment on` $\mathcal{T}_{k,com}$ `******* */`

Update first derivative by $\hat{F}_k^{(1)}(u)$ according to (13).

Update estimate of $\hat{h}_k(v)$ according to (12).

for $t \in \mathcal{T}_{k,com}$ **do**

Predict $\hat{\mu}_k(\mathbf{x}_t)$ according to DNN or TDNN (7).

Set offer price

$$\hat{p}_t = \min\{\max\{\hat{h}_k(\hat{\mu}_k(\mathbf{x}_t)), 0\}, B\},$$

Output: Offered price $\hat{p}_t$, $t \geqslant 1$

Important Result

The finite-sample uniform bound for the DNN and TDNN estimators of $\mathring{\mu}(\mathbf{x})$ are presented below.

1. Suppose density function $f(\cdot)$ of $\mathbf{x}_t$ and the mean utility function $\mathring{\mu}(\cdot)$ are two times continuously differentiable with bounded second-order derivatives on the compact support $\mathcal{X}$ of $\mathbf{x}_t$. The density $f(\mathbf{x})$ is bounded away from 0 and ∞ for $\mathbf{x} \in \mathcal{X}$. Then we have for the DNN estimator that with probability $1 - 2\delta$,

$$\sup_{x \in \mathcal{X}} |\hat{\mu}_k(\mathbf{x}; s) - \mathring{\mu}(\mathbf{x})| \leqslant C s^{-2/d} + B\sqrt{\frac{2s[\log \delta^{-1} + \log d + d \log n]}{n}}. \tag{14}$$

2. Assume density function $f(\cdot)$ of $\mathbf{x}_t$ and the mean utility function $\mathring{\mu}(\cdot)$ are four times continuously differentiable with bounded fourth-order derivatives on the compact support $\mathcal{X}$ of $\mathbf{x}_t$. Then we have for TDNN estimator $D_n(s_1, s_2)(\mathbf{x})$ with $c_1 s_2 \leqslant s_1 \leqslant c_2 s_2$ that with probability $1 - \delta$,

$$\sup_{\mathbf{x} \in \mathcal{X}} |\hat{\mu}_k(\mathbf{x}; s_1, s_2) - \mathring{\mu}(\mathbf{x})| \leqslant C s_1^{-\min\{3, 4/d\}} + CB\sqrt{\frac{2s_1[\log \delta^{-1} + \log d + d \log n]}{n}}. \tag{15}$$

Empirical Analysis

We conduct large-scale simulations to illustrate the performance of our policy. Recall the pricing model is characterized by (1) and (2). Here, the mean utility function is set to be: $\mathring{\mu}(\mathbf{x}_t) = 2(x_{t1} - 1)^2 + 2(x_{t2} - 1)^2 + 2(x_{t3} - 1)^2$, where the contextual covariates $\{\mathbf{x}_t\}_{t \geqslant 1}$ are i.i.d. bounded random vectors of dimension $d = 3$, and consisting of independent entries uniformly distributed according to $(0, 2)$. The noise distribution ϵ_t are i.i.d. noise following $f(z) \propto (1/$

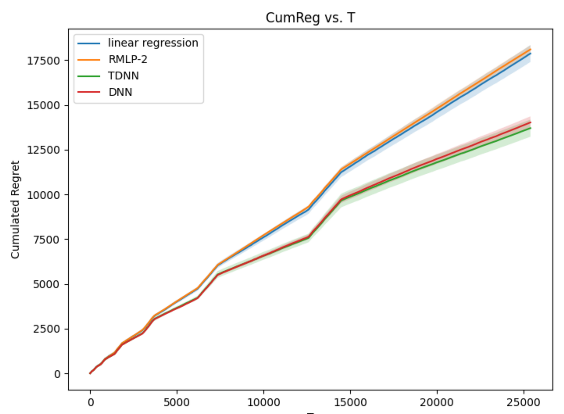

Figure 1:Cumulative Regret against T.

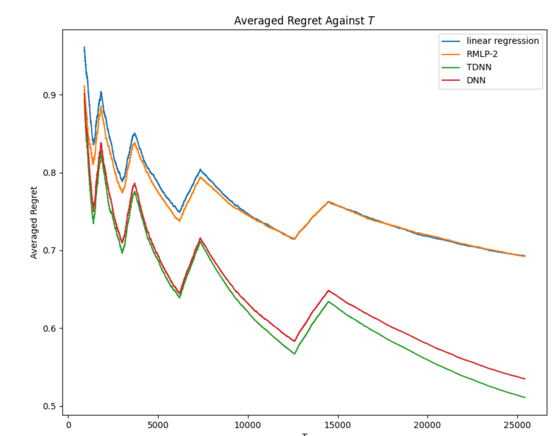

Figure 2:The averaged regret against T, where $\mathrm{AveReg}(T) = \sum_{t=1}^{T} Reg(t)/T$.

We can see from the plots that the regrets we achieved are much smaller than other approaches. As for the comparison with linear regression([2]), TDNN and DNN estimations are robust to the misspecification of the mean utility function and in comparison with 'RMLP-2', our method is robust to the misspecification of the noise distribution since our algorithm can adapt to all functions in the non-parametric class. In contrast, the decay in the averaged regret of exploitation phase for those two benchmark approaches nearly goes to zero. This occurs because the increasing the size of sample points during the exploration phase does not enhance the estimation accuracy of the mean utility function or the noise distribution due to mis-specification.

References

[1] E. Demirkaya, Y. Fan, L. Gao, J. Lv, P. Vossler, and J. Wang.
Optimal nonparametric inference with two-scale distributional nearest neighbors.
Journal of the American Statistical Association, 2022.

[2] J. Fan, Y. Guo, and M. Yu.
Policy optimization using semiparametric models for dynamic pricing.
Journal of the American Statistical Association, 2022.

Contact Information

- Email: jl15681@stern.nyu.edu
- Phone: +1 (929) 672 2671

NEW YORK UNIVERSITY