Perishable Online Inventory Control with Context-Aware Demand Distributions

Yuxiao Wen
New York University
New York, NY 10002
yuxiaowen@nyu.edu

Jingkai Huang
New York University
New York, NY 10002
jh9959@nyu.edu

Weihua Zhou
Zhejiang University
Zhejiang, China 310058
larryzhou@zju.edu.cn

Zhengyuan Zhou
New York University
New York, NY 10002
zzhou@stern.nyu.edu

Abstract

We study the online contextual inventory control problem with perishable goods. We consider a more realistic—and more challenging—setting where the demand depends linearly on observable features (as is standard), but the (residual) noise distribution depends non-parametrically on the features. Surprisingly, little is known when the noise is context-dependent, which captures the heteroskedastic uncertainty in demand that is important in inventory control. Unfortunately, the optimal inventory quantity in this more general setting is no longer a linear function of features (as is the case in the standard setting), making online gradient descent—the gold standard therein—inapplicable. We first present a minimax regret lower bound $\Omega(\sqrt{dT} + T^{\frac{p+1}{p+2}})$, which characterizes the fundamental limit of this learning problem. Here d is the feature dimension, and $p \le d$ is an underlying dimension that captures the intrinsic complexity of the noise distribution. Further, we propose an algorithm achieves the near-optimal regret $\widetilde{O}(\sqrt{dT}+T^{\frac{p+1}{p+2}})$. Additionally, under mild regularity conditions on the noise, we can achieve the improved $\widetilde{O}(\sqrt{dT} + p\sqrt{T})$ regret. To our best knowledge, our results provide the first minimax optimal characterization for online inventory control with context-dependent noise.

1 Introduction

Inventory control under uncertain demand is a central problem in operations management. In many real-world systems, a decision-maker (DM) must repeatedly choose inventory levels over a time horizon, facing random demand and incurring overstocking or understocking costs [9]. A widely used modeling approach assumes that the demand at time t takes the form $D_t = \boldsymbol{\theta}_*^\top \boldsymbol{x}_t + \epsilon_t$, where $\boldsymbol{\theta}_* \in \mathbb{R}^d$ is an unknown parameter, $\boldsymbol{x}_t \in \mathbb{R}^d$ the observable features (context), and ϵ_t an i.i.d. random noise independent of context [2, 3, 5]. However, to derive regret guarantees, previous results crucially rely on that the optimal context-dependent solution is linear in the context under i.i.d. noise, allowing them to compute the loss gradient and apply OSGD. This fails as soon as ϵ_t is not i.i.d.

Additionally, despite its statistical simplicity and interpretability, this linear model can fail in practice. In many applications, the variability of demand depends strongly on the context. For example, in e-commerce platforms, demand uncertainty can vary with user types, geographic regions, or temporal

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

factors such as holidays or promotions. Such heteroskedasticity is well-documented in the empirical inventory literature [8, 6, 7], yet this cannot be captured by the standard homoskedastic model.

To gain more insights on when the heteroskedasticity will occur, consider a natural e-commerce setting where customer-level purchase decisions are modeled as independent Bernoulli events: at time t, the demand arises from n independent customers, each purchasing with probability $p(\boldsymbol{x}_t)$ depending on the context $\boldsymbol{x}_t$. Then the aggregate demand D_t follows a Binomial distribution with mean $np(\boldsymbol{x}_t)$ and variance $np(\boldsymbol{x}_t)(1-p(\boldsymbol{x}_t))$, both of which are context-dependent. For instance, if we consider the sales of umbrellas, when there is zero precipitation, $p(\boldsymbol{x}_t)$ is close to 0 and the demand is almost deterministic. When the precipitation level is intermediate, $p(\boldsymbol{x}_t)$ can be at a constant level, leading to an $O(n)$ variance. Nonetheless, this simple example is not captured by the standard demand model even if $np(\boldsymbol{x}_t)$ is linear in $\boldsymbol{x}_t$, and failing to properly capture the heteroscedasticity can lead to significant additional loss, as highlighted by the empirical studies above.

In this paper, we study online inventory control with a *context-aware* demand distribution, in a general semi-parametric framework. We assume that the demand mean is linear in $\boldsymbol{x}_t \in \mathbb{R}^d$, while the noise distribution may vary with context through a lower-dimensional mapping $\psi(\boldsymbol{x}_t) \in \mathbb{R}^p$. Here $p \le d$ reflects the intrinsic complexity of the distributional dependence of noise on the context. Note $p = 1$ in the Binomial model above. Our primary contributions are:

- We formalize the setting of online contextual inventory control with context-aware demand distributions, and characterize the minimax regret lower bound $\Omega(\sqrt{dT} + T^{\frac{p+1}{p+2}})$.
- We propose an algorithm that achieves the near-optimal regret $\widetilde{O}(\sqrt{dT} + T^{\frac{p+1}{p+2}})$. Under mild regularity conditions on noise, the regret guarantee is improved to $\widetilde{O}(\sqrt{dT} + p\sqrt{T})$.

2 Problem Formulation

Over a time horizon $t \in [T]$, the DM observes a context vector $\boldsymbol{x}_t \in \mathbb{R}^d$ and chooses an inventory level $c_t \in [0, M]$. Then a random demand $D_t \in [0, M]$ is realized, and the DM incurs a loss:

$$\ell_t(c_t) = h\mathbb{E}\Big[(c_t - D_t)^+ \,\big|\boldsymbol{x}_t\Big] + b\mathbb{E}\Big[(D_t - c_t)^+ \,\big|\boldsymbol{x}_t\Big].$$

At the end of each period t, any amount of overstocking perishes, and the DM observes demand realization D_t. The DM's objective is to minimize the cumulative loss over T periods. We assume that the demand satisfies $D_t = \boldsymbol{\theta}_*^\top \boldsymbol{x}_t + \epsilon_t$, where $\boldsymbol{\theta}_* \in \mathbb{R}^d$ is an unknown parameter with $\|\boldsymbol{\theta}_*\|_2 \le 1$, and ϵ_t is a mean-zero noise. Crucially, we allow the distribution of ϵ_t to depend on the context $\boldsymbol{x}_t$: let $Q_t(\cdot) := Q(\cdot; \boldsymbol{z}_t)$ denote the conditional CDF of ϵ_t given the context $\boldsymbol{x}_t$, where $\boldsymbol{z}_t = \psi(\boldsymbol{x}_t) \in \mathbb{R}^p$ is a transformed feature with $p \le d$.

The transformation mapping ψ reflects the DM's *heuristic/prior* on the complexity and contextual dependencies of the noise distributions. In the worst case, the DM can choose $\psi(\boldsymbol{x}) = \boldsymbol{x}$ be the identity to incorporate arbitrary dependence on context. The better heuristic/prior the DM has, the better bound we will obtain, since the transformed dimension p appears in $T^{\frac{p+1}{p+2}}$ in the regret guarantees.

Assumption 1 (Lipschitz CDF). *The noise CDF $Q(u; \psi(\boldsymbol{x}))$ is L-Lipschitz in both u and $\psi(\boldsymbol{x})$.*

Assumption 2 (Transformed features). *The transformed space $\psi(\mathcal{X})$ remains in the unit ball. Also, with $\boldsymbol{x} \sim f_{\boldsymbol{x}}$, the density of $\psi(\boldsymbol{x})$, denoted by f_ψ, is L_ψ-Lipschitz and is upper bounded by $\overline{f}_\psi$.*

Besides the Lipschitzness assumptions above, we also assume stochastic contexts. However, we do not require the density lower bound $f_\psi(\boldsymbol{z}) \ge c > 0$ nor the eigenvalue lower bound on the covariance matrix of $F_{\boldsymbol{x}}$, i.e. $\mathbb{E}_{\boldsymbol{x} \sim F_{\boldsymbol{x}}}[\boldsymbol{x}\boldsymbol{x}^\top] \succeq \lambda \boldsymbol{I}$. These assumptions are common in the literature of online learning with contexts to guarantee uniform convergence over the entire space [1, 3, 4]. In this work, we bypass them by showing convergence along the directions of the *realized* contexts.

Assumption 3 (Stochastic contexts). *The contexts $\boldsymbol{x}_t \in \mathcal{X}$ are generated i.i.d. from an underlying distribution with density $f_{\boldsymbol{x}}$. For simplicity, we assume $\mathcal{X}$ lies in the unit ball, i.e. $\|\boldsymbol{x}_t\|_2 \le 1$.*

To formalize the learning objective, we compete against the optimal time-varying oracle that has full knowledge of $\boldsymbol{\theta}_*$ and Q, i.e. the regret is defined as $\mathsf{Reg}(\pi) := \mathbb{E}\Big[\sum_{t=1}^T \ell_t(c_t) - \ell_t(c_t^*)\Big]$,

Algorithm 1: Contextual online inventory control under context-aware noise

```
1 Input: Time horizon $T$, choice space $\mathcal{C} = [0, M]$, unit costs $b, h > 0$, Lipschitz constant $L > 0$.
2 for $t = 1, 2, \ldots, T$ do
3     Observe the context vector $\boldsymbol{x}_t \in \mathbb{R}^d$.
4     Solve Ridge regression with $\boldsymbol{A}_t \leftarrow \boldsymbol{I} + \sum_{\tau<t} \boldsymbol{x}_\tau \boldsymbol{x}_\tau^\top$ and $\boldsymbol{b}_t \leftarrow \sum_{\tau<t} D_\tau \boldsymbol{x}_\tau$.
5     Compute the estimator $\widehat{\boldsymbol{\theta}}_t \leftarrow \boldsymbol{A}_t^{-1} \boldsymbol{b}_t$.
6     Estimate conditional CDF $\widehat{Q}_t$ from $\{(\epsilon_\tau, \boldsymbol{z}_\tau)\}_{\tau<t}$ and $\boldsymbol{z}_t$ via the NW estimator in (1).
7     Compute the loss estimator $\widehat{\ell}_t(c)$ as in (2) for every $c \in \mathcal{C}$.
8     Order the inventory quantity $c_t \leftarrow \arg\max_{c \in \mathcal{C}} \widehat{\ell}_t(c)$ and observe the realized demand $D_t$.
9     Compute the estimated noise term $\epsilon_t \leftarrow D_t - \boldsymbol{x}_t^\top \widehat{\boldsymbol{\theta}}_t$.
```

where c_t is the inventory decision made by the DM's policy π, and $c_t^* := \arg\min_{c \in [0,M]} \ell_t(c) = \boldsymbol{\theta}_*^\top \boldsymbol{x}_t + Q^{-1}\left(\frac{b}{b+h}; \psi(\boldsymbol{x}_t)\right)$ is the optimal decision for inventory control.

3 Main Results

3.1 Minimax Regret Lower Bound

We start by establishing a regret lower bound that demonstrates the fundamental hardness of contextual inventory learning with context-aware noise. The inherent statistical complexity of this problem naturally arises from two aspects: (1) the estimation of the parametric model $\mathbb{E}_t[D_t] = \boldsymbol{\theta}_*^\top \boldsymbol{x}_t$, and (2) the non-parametric dependence of noise CDF Q on $\psi(\boldsymbol{x}_t)$. We remark that the difficulty of learning does arise from estimating $\boldsymbol{\theta}_*$ and Q, rather than the structure of ψ or $f_{\boldsymbol{x}}$. Indeed, our lower bound construction will only involve a simple linear ψ and a uniform $f_{\boldsymbol{x}}$ over a finite support.

Theorem 1 (Lower bound)**.** *When $T \geq d^2$, it holds that*

$$\inf_{\pi} \sup_{\boldsymbol{\theta}_*, Q, \psi, f_{\boldsymbol{x}}} \mathsf{Reg}(\pi) = \Omega\left((b+h)M\left(\sqrt{dT} + T^{\frac{p+1}{p+2}}\right)\right),$$

where inf is taken over all possible policies, and sup is taken over the problem parameters that satisfy $\|\boldsymbol{\theta}_*\|_2 \leq 1$ and Assumptions 1–3.

3.2 An Algorithm with Matching Upper Bound

Now we introduce and analyze our algorithm for contextual inventory control. At each period t, we estimate Q_t via Kernel regression and $\boldsymbol{\theta}_*$ via Ridge regression respectively. The core of our analysis is to address the mutual dependence between these two estimations and guarantee convergence along "high-probability directions" without strong assumptions on the density.

First, we solve the Ridge regression for $\widehat{\boldsymbol{\theta}}_t$ in Line 6 of Algorithm 1. The following lemma characterizes the performance of our Ridge estimator.

Lemma 3.1. *With probability at least $1 - T^{-2}$, $|\boldsymbol{x}_t^\top \widehat{\boldsymbol{\theta}}_t - \boldsymbol{x}_t^\top \boldsymbol{\theta}_*| \leq (\sqrt{\log(2T)} + 1)\|\boldsymbol{x}_t\|_{\boldsymbol{A}_t^{-1}}$.*

Non-parametric Regression with Measurement Errors To estimate the context-dependent CDF, we propose to use the Nadaraya-Watson (NW) kernel regression and derive its corresponding error bound δ_t: for $u \in \mathcal{C}$ and $\boldsymbol{z}_t = \psi(\boldsymbol{x}_t)$, define

$$\widehat{Q}_t(u; \boldsymbol{z}_t) := \sum_{\tau=1}^{t-1} \frac{K_{a_t}(\boldsymbol{z}_\tau - \boldsymbol{z}_t)\mathbb{1}\left[D_\tau - \boldsymbol{x}_\tau^\top \widehat{\boldsymbol{\theta}}_\tau \leq u\right]}{\sum_{\tau<t} K_{a_t}(\boldsymbol{z}_\tau - \boldsymbol{z}_t)} \tag{1}$$

where K is a smoothing kernel (e.g. Gaussian kernel), $a_t > 0$ the bandwidth parameter, and K_{a_t} is the rescaled kernel. Different from the standard kernel regression, we do not observe the target quantities $\{\epsilon_\tau\}_{\tau<t}$, i.e. the noise realizations. Rather, we only have access to $D_\tau - \boldsymbol{x}_\tau^\top \widehat{\boldsymbol{\theta}}_\tau$ as an

approximate measurement whose error is determined by the performance of $\widehat{\boldsymbol{\theta}}_\tau$. Instead of $\widehat{\boldsymbol{\theta}}_t$, $\widehat{\boldsymbol{\theta}}_\tau$ are used for technical reasons. Let $f_{a_t}(\boldsymbol{z}) = \frac{1}{t-1}\sum_{\tau\in[t-1]} K_{a_t}(\boldsymbol{z}_\tau - \boldsymbol{z})$ be the kernel-smoothed estimator for $f_\psi(\boldsymbol{z})$.

Lemma 3.2. *Under Assumptions 1–3, with probability at least* $1 - T^{-2}$*,*

$$\left|\widehat{Q}_t(u;\boldsymbol{z}) - Q(u;\boldsymbol{z})\right| \le \frac{C_0\log(T)^{\frac{3}{2}}}{f_{a_t}(\boldsymbol{z})}\left(L\sqrt{d/t} + t^{-\frac{1}{p+2}}\right) =: \delta_t(\boldsymbol{z})$$

for every $u \in \mathcal{C}$ *and* $\boldsymbol{z} \in \psi(\mathcal{X})$*, with the constant* C_0 *depends on* K *and* $\overline{f}_\psi$*.*

This conditional bound $\delta_t(\boldsymbol{z})$ is crucial to removing the density lower bound commonly required in the literature (e.g. [4, Assumption 4.2]). Clearly, our estimation is worse and potentially broken for new context $\boldsymbol{z}_t$ with small probability. Yet since we consider the cumulative expected regret, it turns out sufficient to achieve the optimal regret with an error bound scaling inversely with $f_{a_t}(\boldsymbol{z}_t)$. To give a high-level idea, we can show $f_{a_t}(\boldsymbol{z}_t) = \Theta(f_\psi(\boldsymbol{z}_t))$ when $f_\psi(\boldsymbol{z}_t) = \Omega(L\sqrt{d/t} + t^{-\frac{1}{p+2}})$. Consequently, the expected estimation error introduced by (1) at time t can be bounded by

$$\begin{aligned}\mathbb{E}_t[\delta_t] &\le C_0\log(T)^{\frac{3}{2}} O\Big(\mathbb{P}_t\Big(f_\psi(\boldsymbol{z}_t) = o(L\sqrt{d/t} + t^{-\frac{1}{p+2}})\Big) + \mathbb{E}_t\Big[(L\sqrt{d/t} + t^{-\frac{1}{p+2}})/f_\psi(\boldsymbol{z}_t)\Big]\Big)\\ &\le C_0\log(T)^{\frac{3}{2}} O\Big(|\psi(\mathcal{X})|(L\sqrt{d/t} + t^{-\frac{1}{p+2}})\Big).\end{aligned}$$

Finally, based on the estimators $\widehat{Q}_t$ and $\widehat{\boldsymbol{\theta}}_t$, we define a plug-in loss estimator for the conditional expected loss ℓ_t as:

$$\widehat{\ell}_t(c) = h\int_0^c \widehat{Q}_t(y - \widehat{\boldsymbol{\theta}}_t^\top \boldsymbol{x}_t)\mathrm{d}y + b\int_c^M \Big[1 - \widehat{Q}_t(y - \widehat{\boldsymbol{\theta}}_t^\top \boldsymbol{x}_t)\Big]\,\mathrm{d}y. \tag{2}$$

Theorem 2. *Under Assumptions 1–3, with Gaussian kernel* K *and bandwidth* $a_t \asymp t^{-\frac{1}{p+2}}$*,*

$$\mathsf{Reg}(\mathsf{Alg}\ 1) = O\Big((b+h)M(L+1)\log(T)^{\frac{3}{2}}\Big(\sqrt{dT} + T^{\frac{p+1}{p+2}}\Big)\Big) = \widetilde{O}\Big(\sqrt{dT} + T^{\frac{p+1}{p+2}}\Big).$$

3.3 Breaking the Curse of Dimensionality with Benign Distributions

Although the result in Theorem 2 is minimax optimal, it suffers from the curse of dimensionality due to the term $T^{\frac{p+1}{p+2}}$. This section takes inspiration from [4] and provides an arguably mild regularity condition on f_ψ, under which we achieve $\widetilde{O}(\sqrt{dT} + p\sqrt{T})$ regret. Assumption 4 asks the Fourier coefficients of the feature density $f_\psi(\boldsymbol{z})$ and the unconditional probability $Q(u;\boldsymbol{z})f_\psi(\boldsymbol{z})$ (for fixed $u \in [0, M]$) to decay at a fast rate.

Assumption 4. *There exist constants* $c_{FT}, C_{FT}, \omega > 0$ *such that for every* $\boldsymbol{v} \in \mathbb{R}^p$ *and* $u \in \mathcal{C}$*,*

$$\max\{|\mathcal{T}[Q(u;\cdot)f_\psi(\cdot)](\boldsymbol{v})|, |\mathcal{T}[f_\psi](\boldsymbol{v})|\} \le C_{FT}\exp(-c_{FT}\|\boldsymbol{v}\|_2^\omega)$$

where $\mathcal{T}[f](\boldsymbol{v}) = \int_{\mathbb{R}^p} f(\boldsymbol{z})e^{-i\boldsymbol{v}^\top\boldsymbol{z}}\mathrm{d}\boldsymbol{z}$ *denotes the Fourier Transform of* f*.*

Lemma 3.3. *Under Assumptions 1–4, there exists an infinitely smooth kernel* K *such that with probability at least* $1 - T^{-2}$*,*

$$\left|\widehat{Q}(u;\boldsymbol{z}) - Q(u;\boldsymbol{z})\right| \le \frac{\gamma'}{f_{a_t}(\boldsymbol{z})}\left(L\sqrt{d/t} + p/\sqrt{t}\right) =: \delta_t(\boldsymbol{z})$$

for every $u \in \mathcal{C}$ *and* $\boldsymbol{z} \in \psi(\mathcal{X})$*, with* $\gamma' = O\Big(\log(T)^{\frac{p}{\omega}}\log\log(T)^{\frac{1}{2}} + \log(T)^{\frac{3}{2}}\Big)$*.*[1]

Consequently, we arrive at the following improved regret guarantee for Algorithm 1 by replacing Lemma 3.2 with Lemma 3.3 in the analysis:

Theorem 3. *Under Assumptions 1–4, it holds that* $\mathsf{Reg}(\mathsf{Alg}\ 1) = \widetilde{O}\Big(\sqrt{dT} + p\sqrt{T}\Big)$*.*

[1] We remark that this kernel K is problem-independent, while the bandwidth a_t uses the knowledge of the constants c_{FT} and ω in Assumption 4.

References

[1] Ashwinkumar Badanidiyuru, Zhe Feng, and Guru Guruganesh. Learning to bid in contextual first price auctions. In *Proceedings of the ACM Web Conference 2023*, pages 3489–3497, 2023.

[2] Gah-Yi Ban and Cynthia Rudin. The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108, 2019.

[3] Jingying Ding, Woonghee Tim Huh, and Ying Rong. Feature-based inventory control with censored demand. *Manufacturing & Service Operations Management*, 26(3):1157–1172, 2024.

[4] Jianqing Fan, Yongyi Guo, and Mengxin Yu. Policy optimization using semiparametric models for dynamic pricing. *Journal of the American Statistical Association*, 119(545):552–564, 2024.

[5] Jingkai Huang, Kevin Shang, Yi Yang, Weihua Zhou, and Yuan Li. Taylor approximation of inventory policies for one-warehouse, multi-retailer systems with demand feature information. *Management Science*, 71(1):879–897, 2025.

[6] John J Kanet, Michael F Gorman, and Martin Stößlein. Dynamic planned safety stocks in supply networks. *International Journal of Production Research*, 48(22):6859–6880, 2010.

[7] Tatpong Katanyukul, William S Duff, and Edwin KP Chong. Approximate dynamic programming for an inventory problem: Empirical comparison. *Computers & Industrial Engineering*, 60(4): 719–743, 2011.

[8] Xiaolong Zhang. Inventory control under temporal demand heteroscedasticity. *European Journal of Operational Research*, 182(1):127–144, 2007.

[9] Paul H. Zipkin. *Foundations of Inventory Management*. Irwin/McGraw-Hill series in operations and decision sciences. McGraw-Hill, 2000.