**Columbia B9136** April 1st, 2025

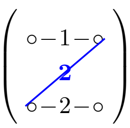
# Lecture 8 — 7 Proofs of Data-driven Newsvendor, part 1

*Lecturer: Will Ma* *Scribe: Jiayu Li and Yunfan Zhang*

## 1 Framework for Decision-Making Under Uncertainty

- $\pi$: decision (often called a "policy").
- $\xi$: uncertainty realization (the "outcome").
- $\pi(\xi)$: reward of policy $\pi$ on outcome $\xi$.
- **Stochastic setting**: $\xi \sim F$, and we know $F$.
- **Adversarial setting**: $\xi$ could be anything.

In lecture 8-9, we consider the case where we see only a *noisy* version of $F$. Examples:

| **Problem** | **Outcome** | **Policy** | **Loss/Reward** | **Typical Framework** | **Benchmark** |
|---|---|---|---|---|---|
| Optimal Stopping | Valuation sequence $\mathbf{v}$ | Non-anticipative stopping policy $\pi$ | e.g. $\pi =$ static threshold (5), then $\pi(\{4,6,8\}) = 6$ | Stochastic | $\sup_\pi \mathbb{E}_{\xi\sim F}\big[\pi(\xi)\big] \le \mathrm{OFF}(F)$ |
| Online Matching | Sequence of arriving vertices $\mathcal{I}$ | Non-anticipative matching policy $\pi$ | e.g. $\pi =$ greedy, then $\pi$  $= 2$ | Adversarial | $\inf_{\mathcal{I}} \dfrac{P(\mathcal{I})}{\mathrm{OFF}(\mathcal{I})}$ |
| Assortment Optimization | List of preferences $\ell$ | Choice of assortment $S$ | e.g. $\pi\big(\{1,2\},3,2,4\big) = r_2$ | Stochastic | $\sup_\pi \mathbb{E}_{\xi\sim F}\big[\pi(\xi)\big]$ |
| Binary Classification | $(x,y)$: (image, label) | Classifier $h$ | $\ell\big(h,(x,y)\big) = \mathbf{1}\{\,h(x) \neq y\,\}$ | Statistical Learning Theory | $\inf_\pi \mathbb{E}_{\xi\sim F}\big[\ell(\pi,\xi)\big]$ |
| Newsvendor | Demand $\xi$ | Stocking quantity $a \in \mathbb{R}$ | $\ell(a,\xi) = b[\xi - a]^+ + h[a - \xi]^+$ | Statistical Learning Theory | $\min_a \mathbb{E}_{\xi\sim F}\big[\ell(a,\xi)\big]$ |

### Statistical Learning Theory Framework

- We observe $N$ i.i.d. samples $(\xi_i)_{i=1}^N$ drawn from $F$, but we do *not* know $F$.
- Let $\ell(\pi,\xi)$ denote the *loss* of policy $\pi$ on outcome $\xi$.
- Define the expected loss of policy $\pi$ under $F$ as

$$L_F(\pi) \;=\; L(\pi) \;=\; \mathbb{E}_{\xi\sim F}\big[\ell(\pi,\xi)\big].$$

- The *optimal policy* (knowing $F$) is given by

$$\pi^* \;\in\; \operatorname{argmin}_\pi L(\pi).$$

- Our goal is to select a policy $\pi$ such that the *regret* (the additive error), defined as the difference $L(\pi) - L(\pi^*)$, vanishes as $N \to \infty$.

### 1.1 Newsvendor Problem

The *newsvendor* problem is a classic single-period inventory model in which we must decide on a stocking quantity $a \in \mathbb{R}$ before observing a random demand $\xi$. After the demand is realized, we incur one of two possible costs:

- A *backlog cost* $b$ for each unit of demand that is not met (when $\xi > a$, stock too little).
- A *holding cost* $h$ for each unit of inventory that remains unsold (when $a > \xi$, stock too much).

Hence, if we let $[\xi - a]^+ = \max\{\xi - a, 0\}$ and $[a - \xi]^+ = \max\{a - \xi, 0\}$, the one-period cost (or loss) is

$$\ell(a, \xi) \;=\; b\,[\xi - a]^+ \;+\; h\,[a - \xi]^+.$$

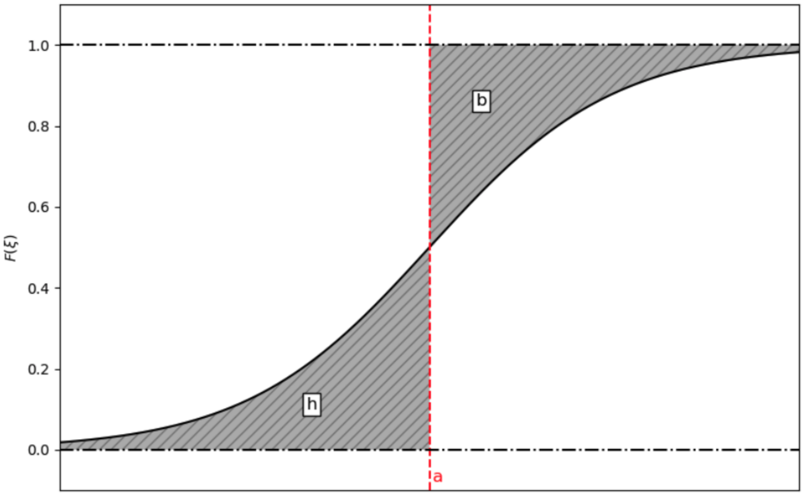


Figure 1: Expected loss decomposition in the newsvendor problem. The shaded area on the left (with box $h$) represents holding cost, while the shaded area on the right (with box $b$) represents backlog cost.

We assume $\xi$ has distribution $F$, then the expected loss if we stock $a$ is

$$\begin{aligned}
L(a) &= \int_\xi \left(b[\xi - a]^+ + h[a - \xi]^+\right) dF(\xi) && (1)\\
&= h\int_0^a F(\xi)\, d\xi + b\int_a^\infty (1 - F(\xi))\, d\xi && (2)\\
&= (h + b)\left(\int_0^a (1 - q)F(\xi)\, d\xi + \int_a^\infty q(1 - F(\xi))\, d\xi\right), \quad \text{where } q = \frac{b}{b + h} && (3)
\end{aligned}$$

The expected loss function defined in equation (1) is visually represented in Figure 1. To arrive at equation (2), we apply integration by parts. Specifically, observe that $\int b[\xi - a]^+ \, dF(\xi) = b\int_a^\infty (\xi - a)\, dF(\xi)$, then rewrite this integral as:

$$b\int_a^\infty (\xi - a)\, dF(\xi) = b\int_a^\infty \int_a^\xi 1\, dt\, dF(\xi) = b\int_a^\infty \Big[\int_t^\infty dF(\xi)\Big] dt = b\int_a^\infty \big(1 - F(t)\big)\, dt.$$

We can apply the same reasoning to the holding cost term,

$$h\int [a-\xi]^+ \, dF(\xi) = h\int_0^a (a-\xi)\, dF(\xi) = h\int_0^a \int_\xi^a 1\, dt\, dF(\xi) = h\int_0^a \Big[\int_0^t dF(\xi)\Big] dt = h\int_0^a F(t)\, dt,$$

thus, we arrive at equation (2). From (3), the term $\int_0^a (1-q)F(\xi)\, d\xi$ suggests that for $\xi < a$, we want to choose $a$ so that $F(\xi) \le q$, while the term $\int_a^\infty q(1 - F(\xi))\, d\xi$ suggests that for $\xi > a$, we want to choose $a$ so that $F(\xi) \ge q$. Hence, the optimal solution satisfies - when the distribution $F$ is continuous,

$$a^* \;=\; F^{-1}(q),$$

Similarly, as illustrated by Figure 2, when the demand distribution $F$ is discrete, the optimal stock is chosen by

$$a^* \;=\; \inf\big\{\, a : F(a) \;\ge\; q \big\} \quad \text{where} \quad q \;=\; \frac{b}{b+h}.$$

Specifically, there are two scenarios:

- *Left panel (unique optimum):* The cumulative distribution $F$ crosses the horizontal line at $q$ precisely at one integer value. Hence, there is a *single* stocking level $a^*$ for which $F(a^*) \ge q$ (and $F(a^* - \epsilon) < q, \forall \epsilon > 0$).
- *Right panel (multiple optima):* The distribution $F$ jumps from below $q$ to above $q$ in a single "step". In that "jump", all values of $a$ are equally optimal, and by convention, we *take the smallest* $a$ in that jump as $a^*$.

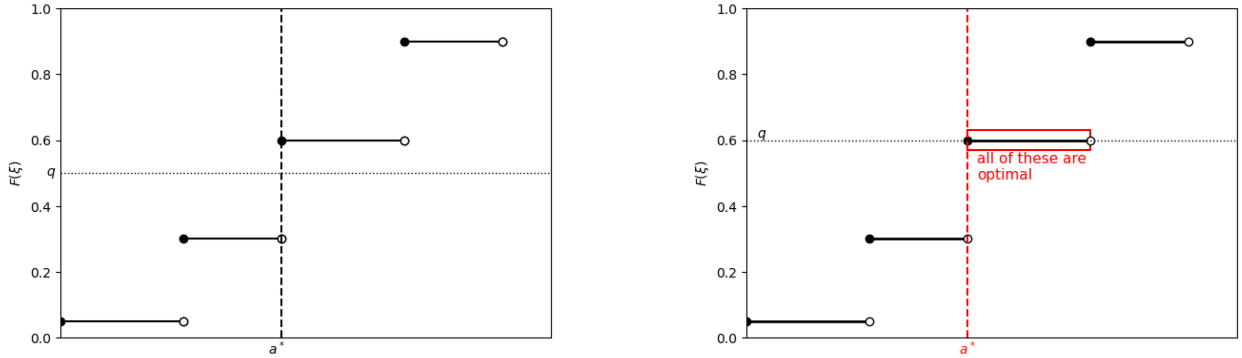


Figure 2: Two discrete distribution settings in the newsvendor problem.

**Note:** To normalize the regret, we assume $h + b = 1$ and $\xi \in [0, 1]$ throughout the proofs.

## 1.2 Proof (1): Direct (DKW Inequality)

Let $A$ denote our decision that depends on the $N$ samples $\xi_1, \ldots, \xi_N$. As illustrated by Figure 3 in discrete setting, define the *regret*:

$$\text{Regret} = L(A) - L(a^*) \; = \; \int_{a^*}^{A} \big(F(\xi) - q\big)\, d\xi.$$

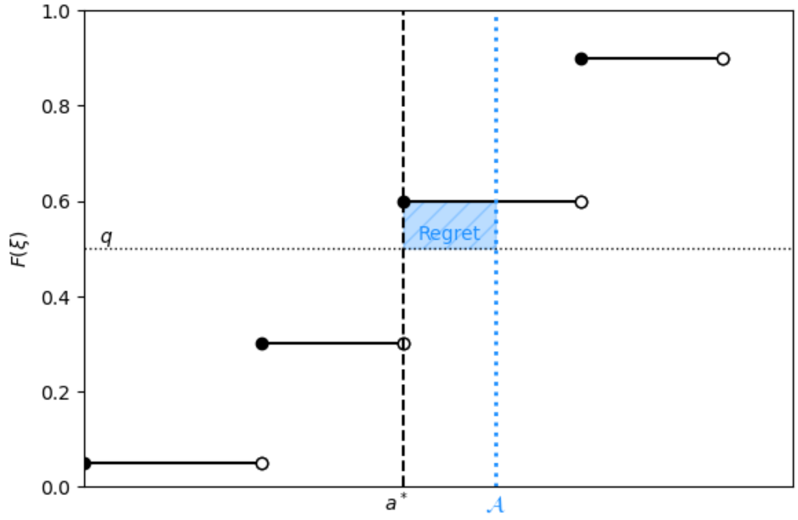


Figure 3: Illustration of regret in the newsvendor problem. The decision $A$ based on the empirical CDF $\hat{F}$ exceeds the optimal $a^*$, resulting in regret quantified by the shaded rectangular area between $a^*$ and $A$, with height $F(\xi) - q$.

In practice, we pick $A$ by constructing the *empirical CDF*:

$$\hat{F}(\xi) \; = \; \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{\xi_i \leq \xi\},$$

and then define the decision

$$\hat{a} \; := \; \inf\{\, a : \hat{F}(a) \; \geq \; q\},$$

where $\hat{a}$ is known as the *Sample Average Approximation* (SAA) in stochastic optimization, or the *Empirical Risk Minimizer* (ERM) in statistical learning problem.

In order to prove the expected regret bound, we need the following claim:

**Claim 1.**

$$L(\hat{a}) - L(a^*) \; \leq \; \sup_a \big| F(a) - \hat{F}(a) \big|.$$

[Will: Note: no need for absolute values on LHS, because $L(\hat{a}) - L(a^*)$ is always non-negative. Can you please change this throughout the note?]

**Proof.** Consider two cases:

**Case 1:** $\hat{a} \leq a^*$. Then

$$\begin{aligned} L(\hat{a}) - L(a^*) &= \int_{\hat{a}}^{a^*} \big(q - F(\xi)\big)\, d\xi && (4) \\ &\leq\ (a^* - \hat{a})\,(q - F(\hat{a})) && (5) \\ &=\ (a^* - \hat{a})\,\Big(q - \hat{F}(\hat{a}) + \hat{F}(\hat{a}) - F(\hat{a})\Big) \\ &\leq\ (a^* - \hat{a})\,\Big(\hat{F}(\hat{a}) - F(\hat{a})\Big) && (6) \\ &\leq\ |F(\hat{a}) - \hat{F}(\hat{a})| && (7) \\ L(\hat{a}) - L(a^*) &\leq\ \sup_a |F(a) - \hat{F}(a)|, && (8) \end{aligned}$$

where (4) holds by definition, (5) follows from the fact that $F(\xi) \geq F(\hat{a})$ for any $\xi \geq \hat{a}$, (6) holds by the definition of $\hat{a}$ which implies that $q - \hat{F}(\hat{a}) \leq 0$, (7) follows from the fact that $a^* - \hat{a} \leq 1$ since $\xi \in [0,1]$; (8) follows from taking the supremum over all $a$.

**Case 2:** $\hat{a} > a^*$. Then

$$\begin{aligned} L(\hat{a}) - L(a^*) &= \int_{a^*}^{\hat{a}} \big(F(\xi) - q\big)\, d\xi \\ &\leq \lim_{a \to \hat{a}_-} (\hat{a} - a^*)\,\Big(F(a) - \hat{F}(a) + \hat{F}(a) - q\Big) && (9) \\ &\leq \lim_{a \to \hat{a}_-} \Big(F(a) - \hat{F}(a)\Big) && (10) \\ &\leq \sup_a |F(a) - \hat{F}(a)|, && (11) \end{aligned}$$

where (9) uses the fact that the integrand $F(\xi) - q$ is bounded from above by its value at the left-limit $\hat{a}_-$; (10) holds because $\xi \in [0,1]$ implying $\hat{a} - a^* \leq 1$, and $\hat{F}(a) - q < 0$ for all $a < \hat{a}$ by definition; thus, the regret is again bounded by the uniform deviation $\sup_a |F(a) - \hat{F}(a)|$ in (11). $\square$

**We also need the DKW Inequality:** We already know that the Hoeffding's Inequality says that for an i.i.d. sample of size $N$,

$$\mathbb{P}\big[\ |\hat{F}(a) - F(a)|\ \geq\ \varepsilon\big]\ \leq\ 2\exp\big(-2N\varepsilon^2\big).$$

DKW Inequality upgrads to a uniform version:

$$\mathbb{P}\Big[\ \sup_a |\hat{F}(a) - F(a)|\ \geq\ \varepsilon\Big]\ \leq\ 2\ \exp\big(-2N\varepsilon^2\big).$$

**We then conclude our proof:** By Claim 1, if $\sup_a |\hat{F}(a) - F(a)| \leq \varepsilon$, then $L(\hat{a}) - L(a^*) \leq \varepsilon$. Consequently, applying the DKW Inequality, we have

$$\mathbb{P}\big[\ L(\hat{a}) - L(a^*) \leq \varepsilon\big]\ \geq\ 1 - 2\ \exp\big(-2N\varepsilon^2\big).$$

Hence,

$$
\begin{aligned}
\mathbb{E}\left[\, L(\hat{a}) - L(a^*)\,\right] &= \int_0^\infty \mathbb{P}\left(L(\hat{a}) - L(a^*) > \varepsilon\right)\, d\varepsilon \\
&\le \frac{1}{\sqrt{N}} + 2\int_{1/\sqrt{N}}^\infty \exp\left(-2N\varepsilon^2\right)\, d\varepsilon \\
&\le \frac{1}{\sqrt{N}} + 2\int_{1/\sqrt{N}}^\infty \exp\left(-2\sqrt{N}\varepsilon\right)\, d\varepsilon \\
&= \frac{1}{\sqrt{N}} + \frac{2}{2\sqrt{N}} \exp\left(-2\sqrt{N}\varepsilon\right)\Big|_{\infty}^{\varepsilon=1/\sqrt{N}} \\
&= \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{N}} e^{-2} \\
&= \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \qquad (12)
\end{aligned}
$$

where the first inequality follows from the DKW Inequality and the fact that $\mathbb{P}\left(L(\hat{a}) - L(a^*) > \varepsilon\right) \le 1$ on the interval $\left[0, \frac{1}{\sqrt{N}}\right]$; the second inequality holds because $\varepsilon \ge \frac{1}{\sqrt{N}}$ on the interval $\left[\frac{1}{\sqrt{N}}, \infty\right)$.

**Remark 2.** Note that here it is sufficient to integrate up to 1, since $L(\hat{a}) - L(a^*) \le 1$. However, we integrated up to $\infty$ to demonstrate the generality of this trick. The challenge with unbounded demands actually lies in getting a good upper bound on $\mathbb{P}\left(L(\hat{a}) - L(a^*) > \varepsilon\right)$ for large $\varepsilon$.

## 2 Four Approaches to Bound the Expected Regret

We then show four proofs that $\mathbb{E}\big[L(A) - L(a^*)\big] \to 0$ as $N$ grows, all of which start with the decomposition:

$$
\mathbb{E}\big[L(A) - L(a^*)\big] \;=\; \mathbb{E}\big[L(A) - \hat{L}(A)\big] + \mathbb{E}\big[\hat{L}(A) - \hat{L}(a^*)\big] \;+\; \mathbb{E}\big[\hat{L}(a^*) - L(a^*)\big]. \qquad (13)
$$

where

$$
\hat{L}(a) \;:=\; \frac{1}{N}\sum_{i=1}^N \left(b[\xi_i - a]^+ + h[a - \xi_i]^+\right) \;=\; \frac{1}{N}\sum_{i=1}^N \ell(a, \xi_i)
$$

**Remark 3.** Note that $\mathbb{E}\big[\hat{L}(a^*) - L(a^*)\big] = 0$, as $a^*$ is a constant; $\hat{L}(A) - \hat{L}(a^*) \le 0$ if $A = \hat{a}$ is the empirical minimizer; and by taking the supremum, $L(A) - \hat{L}(A) \;\le\; \sup_{a\in[0,1]}(L(a) - \hat{L}(a))$.

### 2.1 Proof (2): Finite Hypothesis Class + Covering Number

Let $\mathcal{A}$ be a finite subset of $[0,1]$, then a union bound combined with Hoeffding's inequality implies

$$
\mathbb{P}\Big[\sup_{a\in\mathcal{A}} \big| L(a) - \hat{L}(a)\big| \;>\; \varepsilon\Big] \;\le\; |\mathcal{A}|\, \exp\big(-2N\varepsilon^2\big). \qquad (14)
$$

More specifically, let $\mathcal{A} = \{0, \frac{1}{M}, \frac{2}{M}, \ldots, 1\}$. Then, consider the following algorithm: choose $A$ in $\mathcal{A}$ that is closest to the true empirical minimizer $\hat{a}$. Under this algorithm, we have

$$
\begin{aligned}
\hat{L}(A) - \hat{L}(\hat{a}) &= \frac{1}{N}\sum_{i=1}^N \left(\ell(A, \xi_i) - \ell(\hat{a}, \xi_i)\right) \\
&\le \frac{1}{N}\sum_{i=1}^N |A - \hat{a}| = |A - \hat{a}| \le \frac{1}{2M}
\end{aligned} \qquad (15)
$$

The first inequality in (15) holds because the loss function $\ell(\cdot, \xi)$ is $\max\{b, h\}$-Lipschitz in its first argument for all $\xi$. That is, for any $\xi$ and any $a, a' \in \mathbb{R}_{\geq 0}$, we have

$$|\ell(a, \xi) - \ell(a', \xi)| \leq \max\{b, h\} \cdot |a - a'|.$$

Applying this Lipschitz bound pointwise gives:

$$|\ell(A, \xi_i) - \ell(\hat{a}, \xi_i)| \leq \max\{b, h\} \cdot |A - \hat{a}| \leq |A - \hat{a}|.$$

The second inequality in (15) follows from the construction of $A$: it is chosen as the closest point to $\hat{a}$ in the grid $\mathcal{A} = \{0, \frac{1}{M}, \ldots, 1\}$, whose spacing is $1/M$. Therefore, the maximum distance between $A$ and $\hat{a}$ is at most half the grid width, i.e., $|A - \hat{a}| \leq \frac{1}{2M}$.
Hence, by (15) and the fact that $\hat{L}(\hat{a}) - \hat{L}(a^*) \leq 0$ as $\hat{a}$ is the empirical minimizer, we obtain

$$\hat{L}(A) - \hat{L}(a^*) = \hat{L}(A) - \hat{L}(\hat{a}) + \hat{L}(\hat{a}) - \hat{L}(a^*) \leq \hat{L}(A) - \hat{L}(\hat{a}) \leq \frac{1}{2M}. \tag{16}$$

Additionally, since $L(A) - \hat{L}(A) \leq \sup_{a \in \mathcal{A}} \Big(L(a) - \hat{L}(a)\Big)$, applying (14) yields:

$$\mathbb{P}\left[L(A) - \hat{L}(A) > \varepsilon\right] \leq (M + 1)\exp(-2N\varepsilon^2).$$

Let $\varepsilon = \sqrt{\frac{\log N}{N}}$, then $\mathbb{P}\left[L(A) - \hat{L}(A) > \varepsilon\right] \leq \frac{M+1}{N^2}$. Hence,

$$\mathbb{E}\left[L(A) - \hat{L}(A)\right] \leq \sqrt{\frac{\log N}{N}}\left(1 - \frac{M+1}{N^2}\right) + \frac{M+1}{N^2}. \tag{17}$$

Combining (16) and (17), and then by setting $M = N$, we obtain

$$\begin{aligned}
\mathbb{E}\left[L(A) - \hat{L}(a^*)\right] &= \mathbb{E}\left[L(A) - \hat{L}(A)\right] + \mathbb{E}\left[\hat{L}(A) - \hat{L}(a^*)\right] \\
&\leq \frac{1}{2N} + \sqrt{\frac{\log N}{N}} + \frac{N+1}{N^2} \\
&= \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)
\end{aligned}$$

## 2.2 Proof (3): Based on VC Theory

Let $A = \hat{a}$ be the empirical risk minimizer. Then, the expected regret (13) is upper bounded by the uniform error:

$$\mathbb{E}[L(A) - L(a^*)] \leq \mathbb{E}\left[\sup_{a \in [0,1]} (L(a) - \hat{L}(a))\right].$$

**Theorem 4** (from learning theory)**.**

$$\mathbb{E}\Big[\sup_{a \in [0,1]} L(a) - \hat{L}(a)\Big] \;=\; \mathcal{O}\Big(\sqrt{\frac{d \cdot \log(N/d)}{N}}\Big),$$

*where $d$ is the pseudo-dimension. This comes from Rademacher complexity and Sauer's lemma. In fact, by a chaining argument, the bound can be refined to*

$$\mathbb{E}\Big[\sup_{a \in [0,1]} L(a) - \hat{L}(a)\Big] \;=\; \mathcal{O}\Big(\sqrt{\frac{d}{N}}\Big).$$

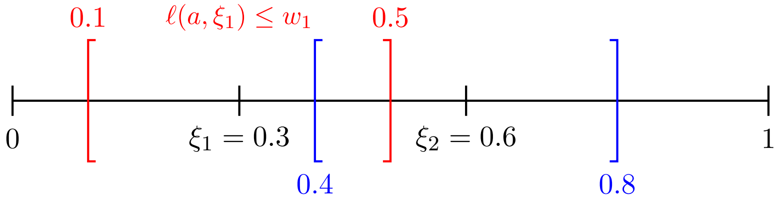
**Definition 5** (Pseudo-dimension)**.** Samples $\xi_1, \ldots, \xi_d$ with witness $w_1, \ldots, w_d$ are *pseudo-shattered* if for all $G \subseteq \{1, \ldots, d\}$ ("Good"), $\exists a \in [0,1]$ such that

$$\begin{aligned} \ell(a, \xi_i) &\leq w_i, \quad \forall i \in G, \\ \ell(a, \xi_i) &> w_i, \quad \forall i \notin G. \end{aligned}$$

That is, action $a$ achieves a "Good" loss that is below the threshold $w_i$ for $i \in G$, and achieves a "Bad" loss that is above the threshold $w_i$ for $i \notin G$. *Pseudo-dimension* is the largest integer $d$ for which one can construct $\xi_1, \ldots, \xi_d$, $w_1, \ldots, w_d$ to be pseudo-shattered.

To show that pseudo-dimension $d$ is finite, it suffices to demonstrate that an infinite set of points cannot be pseudo-shattered. In the following example, we illustrate how two points can be pseudo-shattered. Later, we will prove that $d$ is at most 2.

**Example 6.** Consider $q = \frac{1}{2}$, $b = h = \frac{1}{2}$. One can try to see if we can shatter $d = 2$. For instance, pick $\xi_1 = 0.3$, $\xi_2 = 0.6$, and $w_1 = w_2 = 0.1$. We can place the threshold $a$ in various positions to make $\ell(a, \xi_i)$ either $\leq w_i$ or $> w_i$. This works out for $d = 2$.

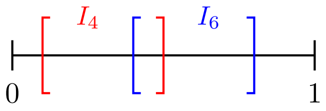


Specifically,

- For $G = \emptyset$ (i.e., $\ell(a, \xi_i) > w_i$ for $i = 1, 2$), choose $a \in [0, 0.1) \cup (0.8, 1]$
- For $G = \{1\}$ (i.e., $\ell(a, \xi_1) \leq w_1,\ \ell(a, \xi_2) > w_2$ ), choose $a \in [0.1, 0.4)$
- For $G = \{2\}$ (i.e., $\ell(a, \xi_2) \leq w_2,\ \ell(a, \xi_1) > w_1$ ), choose $a \in (0.5, 0.8]$
- For $G = \{1, 2\}$ (i.e., $\ell(a, \xi_i) \leq w_i$ for $i = 1, 2$), choose $a \in [0.4, 0.5]$

**Lemma 7.** *$d \leq 2$.*

*Proof.* For each $i$, the set $\{\, a : \ell(a, \xi_i) \leq w_i\}$ is an *interval* $I_i \subseteq [0,1]$.



As $a$ moves continuously from 0 to 1, the indicator $1\{\, a \in I_i\}$ can only change from 0 to 1 or 1 to 0 once or twice, so overall the vector $\left(1\{\, a \in I_i\}\right)_{i=1}^d$ can change at most $2d$ times. Therefore, the set of distinct labelings we can realize is at most $1 + 2d$, i.e.,

$$\left| \left\{ \left(1\{\, a \in I_i\}\right)_{i=1}^d : a \in [0,1] \right\} \right| \leq 1 + 2d.$$

But to shatter $d$ points, we would need $2^d$ labelings. Thus we require $2^d \leq 1 + 2d$. The only integer $d$ satisfying this is $d \leq 2$. $\square$

**Tangent: Optimal Stopping.** We go on a tangent beyond the Newsvendor problem, into the Optimal Stopping problem, to further illustrate pseudo-dimension. We show that *adaptive thresholds policies* $(\tau_t)_{t=1}^T$ can shatter $T$ points.

Adaptive thresholds policies refer to a sequence of thresholds $(\tau_t)_{t=1}^T$, where at each time $t$, we stop if we haven't stopped yet and the current observation exceeds the threshold $\tau_t$. To show that this class of policies can pseudo-shatter $d = T$ points, consider the following construction. For each $t \in [T]$, define $\xi_t = (0, \dots, 0, 0.5, 0, \dots, 0)$ (with the 0.5 in the $t$'th coordinate) and $w_t = 0.25$. Given $G \subseteq [T]$, define the adaptive threshold policy:

$$\tau_t \;=\; \begin{cases} 1, & \text{if } t \notin G, \\ 0.2, & \text{if } t \in G. \end{cases}$$

- For $t \in G$, the threshold at time $t$ is 0.2. Since $\xi_t^{(t)} = 0.5 > 0.2$, the policy will stop at time $t$, and the loss will be small (i.e., $\ell(\tau, \xi_t) \le w_t$).
- For $t \notin G$, the threshold is set to 1. Since $\xi_t^{(t)} = 0.5 < 1$, the policy will not stop at time $t$, and will instead continue past all zero entries, resulting in a large loss (i.e., $\ell(\tau, \xi_t) > w_t$).

In this way, the policy can match any labeling of the inputs by selectively adjusting the thresholds. Hence, the class of adaptive threshold policies can pseudo-shatter $T$ points.

In contrast, if we restrict ourselves to *static threshold policies* ($\tau_t = \tau,\ \forall t \in [T]$), then the pseudo-dimension does *not* grow with $T$ (Homework problem).

## 2.3 Proof (4): Stability

Given data $\boldsymbol{\xi} = (\xi_i)_{i=1}^N$, we revisit the expected loss of any policy $A$ that can be defined as a mapping from $\boldsymbol{\xi}$ to a decision. We write $A(\boldsymbol{\xi})$ as needed to emphasize this dependence, sometimes shortening to $A$ when the data being input is clear. We rewrite $\mathbb{E}[L(A)]$ and $\mathbb{E}[\hat{L}(A)]$ as follows.

$$\mathbb{E}[L(A)] \;=\; \mathbb{E}_{\boldsymbol{\xi},\tilde{\xi}}\big[\ell\big(A(\boldsymbol{\xi}),\, \tilde{\xi}\big)\big] \;=\; \mathbb{E}\big[\ell\big(A(\tilde{\xi}, \xi_2, ..., \xi_N),\, \xi_1\big)\big],$$

To explain, the first equality follows by the definition of expected loss, and the second follows by the symmetry of the $N + 1$ i.i.d. samples $\tilde{\xi}, \xi_1, ..., \xi_N$. Since all of these are identically distributed, we can switch $\tilde{\xi}$ with any $\xi_i$, but for convenience we assume $i = 1$. In particular, the algorithm does not know $\xi_1$.

$$\mathbb{E}[\hat{L}(A)] \;=\; \mathbb{E}_{\boldsymbol{\xi}}\big[\frac{1}{N}\sum_{i=1}^N \ell\big(A(\boldsymbol{\xi}),\, \xi_i\big)\big] \;=\; \mathbb{E}\big[\ell\big(A(\xi_1, \xi_2, ..., \xi_N),\, \xi_1\big)\big],$$

As before, by the symmetry of the $N$ samples, we can replace any $\xi_i$ by $\xi_1$ in the middle expectation. The key difference here is that in the final expectation, the algorithm receives the evaluation sample $\xi_1$ as one of its data points.

Define $\tilde{a} = A(\tilde{\xi}, \xi_2, ..., \xi_N)$ and shorten $A(\boldsymbol{\xi})$ to $A$. Then, we have

$$\begin{aligned} \mathbb{E}[L(A) - \hat{L}(A)] &= \mathbb{E}_{\boldsymbol{\xi},\tilde{\xi}}\Big[\ell\big(A(\tilde{\xi}, \xi_2, ..., \xi_N), \xi_1\big) - \ell\big(A(\boldsymbol{\xi}), \xi_1\big)\Big] && (18) \\ &\leq \sup_{\boldsymbol{\xi},\tilde{\xi}}\Big(\ell\big(A(\tilde{\xi}, \xi_2, ..., \xi_N), \xi_1\big) - \ell\big(A(\boldsymbol{\xi}), \xi_1\big)\Big) \\ &= \sup_{\boldsymbol{\xi},\tilde{\xi}}\Big(\ell\big(\tilde{a}, \xi_1\big) - \ell\big(A, \xi_1\big)\Big). && (19) \end{aligned}$$

**Regularize:**
Define the regularized empirical loss as $\hat{L}^{\text{Reg}}(a) := \hat{L}(a) + \frac{\lambda}{2}a^2$. Let $A$ be the minimizer of $\hat{L}^{\text{Reg}}$:

$$A \ = \ \text{argmin}_{a\in[0,1]} \ \hat{L}^{\text{Reg}}(a).$$

Meanwhile, define $\tilde{L}^{\text{Reg}}(a) := \hat{L}^{\text{Reg}}(a) + \frac{\ell(a,\tilde{\xi})-\ell(a,\xi_1)}{N}$. Let $\tilde{a}$ be the minimizer of $\tilde{L}^{\text{Reg}}$, which is the same mapping as $A$ except on the data $\tilde{\xi}, \xi_2, \ldots, \xi_N$ instead of $\xi_1, \xi_2, \ldots, \xi_N$:

$$\tilde{a} \ = \ \text{argmin}_{a\in[0,1]} \ \tilde{L}^{\text{Reg}}(a).$$

It can be checked that $\hat{L}^{\text{Reg}}(\cdot)$ is a $\lambda$-strongly convex function, which is formally defined as

$$\forall a \in [0,1], \ \exists \text{ a ``subgradient'' } \nabla f(a), \text{ s.t. } \forall \tilde{a} \in [0,1], \ f(\tilde{a}) \geq f(a) + \nabla f(a)\left(\tilde{a} - a\right) + \frac{\lambda}{2}\left(\tilde{a} - a\right)^2.$$

We now have all the ingredients needed to proceed. We obtain

$$\begin{aligned} \hat{L}^{\text{Reg}}(\tilde{a}) - \hat{L}^{\text{Reg}}(A) &= \hat{L}(\tilde{a}) + \frac{\lambda}{2}\tilde{a}^2 - \hat{L}(A) - \frac{\lambda}{2}A^2 \\ &= \tilde{L}(\tilde{a}) + \frac{\lambda}{2}\tilde{a}^2 - \tilde{L}(A) - \frac{\lambda}{2}A^2 + \frac{\ell(\tilde{a}, \xi_1) - \ell(\tilde{a}, \tilde{\xi})}{N} - \frac{\ell(A, \xi_1) - \ell(A, \tilde{\xi})}{N} \\ &\leq \frac{2}{N}|\tilde{a} - A| \end{aligned}$$

The last inequality holds because, by the definition of $\tilde{a}$, we have $\tilde{L}(\tilde{a}) + \frac{\lambda}{2}\tilde{a}^2 - \tilde{L}(A) - \frac{\lambda}{2}A^2 \leq 0$, and by the Lipschitz condition, we have $|\ell(\tilde{a}, \xi_1) - \ell(A, \xi_1)| \leq \max\{b, h\}|\tilde{a} - A| \leq |\tilde{a} - A|$, $|\ell(\tilde{a}, \tilde{\xi}) - \ell(A, \tilde{\xi})| \leq |\tilde{a} - A|$.

Since $A$ minimizes $f(a) = \hat{L}^{\text{Reg}}(a)$, the first-order optimality condition implies $\nabla f(A)(\tilde{a} - A) \geq 0$. Using the strong convexity of $\hat{L}^{\text{Reg}}(a)$, we have

$$\begin{aligned} &\frac{\lambda}{2}(\tilde{a} - A)^2 \leq \hat{L}^{\text{Reg}}(\tilde{a}) - \hat{L}^{\text{Reg}}(A) \\ &\Rightarrow \frac{\lambda}{2}(\tilde{a} - A)^2 \leq \frac{2}{N}|\tilde{a} - A| \\ &\Rightarrow |\tilde{a} - A| \leq \frac{4}{N\lambda}. && (20) \end{aligned}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}[L(A) - L(a^*)] &= \mathbb{E}[L(A) - \hat{L}(A) + \hat{L}(A) - \hat{L}(a^*)] \\ &\leq \frac{4}{N\lambda} + \mathbb{E}[\hat{L}(A) - \hat{L}(\hat{a})] \\ &\leq \frac{4}{N\lambda} + \frac{\lambda}{2} \\ &= \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \end{aligned}$$

The first inequality follows from the bounds in (19) and (20), together with the fact that $\hat{L}(\hat{a}) \leq \hat{L}(a^*)$. The second inequality uses the definition of $A$ as the minimizer of the regularized loss: $\hat{L}(A) + \frac{\lambda}{2}A^2 \leq \hat{L}(\hat{a}) + \frac{\lambda}{2}\hat{a}^2$, which implies $\hat{L}(A) - \hat{L}(\hat{a}) \leq \frac{\lambda}{2} \cdot 1$. The last equality holds by setting $\lambda = \sqrt{\frac{8}{N}}$ to obtain the optimal rate.

## 2.4 Proof (5): On-Average Stability Without Regularization

Recall from (18) that

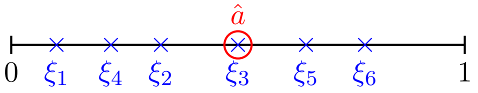
$$\mathbb{E}[L(A) - \hat{L}(A)] \;=\; \mathbb{E}_{\boldsymbol{\xi},\tilde{\xi}}\Big[\ell\big(A(\tilde{\xi}, \xi_2, ..., \xi_N), \xi_1\big) - \ell\big(A(\boldsymbol{\xi}), \xi_1\big)\Big].$$

Letting $A(\cdot)$ be the empirical risk minimization mapping, and letting $\tilde{a} := A(\tilde{\xi}, \xi_2, ..., \xi_N), \hat{a} := A(\boldsymbol{\xi})$, we can derive

$$\begin{aligned}
\mathbb{E}[L(A) - L(a^*)] &\leq \mathbb{E}_{\boldsymbol{\xi},\tilde{\xi}}\big[\,\ell\big(\tilde{a}, \xi_1\big) - \ell\big(\hat{a}, \xi_1\big)\big] \\
&\leq \mathbb{E}_{\boldsymbol{\xi}}\left[\sup_{\tilde{\xi}} |\tilde{a} - \hat{a}|\right] \\
&\leq \mathbb{E}\left[\max\left\{\xi^{(\lfloor qN \rfloor+2)} - \xi^{(\lfloor qN \rfloor+1)}, \xi^{(\lfloor qN \rfloor+1)} - \xi^{(\lfloor qN \rfloor)}\right\}\right].
\end{aligned}$$

The second inequality holds because $\ell(\cdot, \xi_1)$ is Lipschitz. In the final inequality, note that $\hat{a}$ takes the $(\lfloor qN \rfloor+1)$'st smallest number in $\{\xi_1, ..., \xi_N\}$ (assuming $q < 1$), and we let $\xi^{(\ell)}$ denote the $\ell$'th smallest number in $\{\xi_1, ..., \xi_N\}$ with $\xi^{(0)} = 0, \xi^{(N+1)} = 1$. Given $\hat{a} = \xi^{(\lfloor qN \rfloor+1)}$, we claim that the furthest $\tilde{a}$ can deviate is to $\xi^{(\lfloor qN \rfloor+2)}$ or $\xi^{(\lfloor qN \rfloor)}$, which we illustrate through the following example.

For example, let $N = 6$, and suppose $\hat{a}$ is the 4th smallest value in $\{\xi_1, ..., \xi_6\}$, while $\tilde{a}$ is the fourth smallest in $\{\tilde{\xi}, \xi_2, ..., \xi_6\}$.



Then, $\hat{a} = \xi^{(4)} = \xi_3$, and $\tilde{a} = \begin{cases} \xi_5, & \text{if } \xi_5 \leq \tilde{\xi} \leq 1, \\ \xi_3, & \text{if } 0 \leq \tilde{\xi} \leq \xi_3, \\ \tilde{\xi}, & \text{if } \xi_3 \leq \tilde{\xi} \leq \xi_5. \end{cases}$

Since $\xi_3 = \xi^{(4)}$, $\xi_5 = \xi^{(5)}$, we have $|\tilde{a} - \hat{a}| \leq \xi^{(5)} - \xi^{(4)}$. Moreover, this argument holds for any $i \in [N]$; that is, we can switch any $\xi_i$ with $\tilde{\xi}$. Repeating the same analysis:

- If we switch $\tilde{\xi}$ with any of $\xi_1, \xi_4, \xi_2$, we will get $|\tilde{a} - \hat{a}| \leq \xi^{(5)} - \xi^{(4)}$.
- If we switch $\tilde{\xi}$ with any of $\xi_5, \xi_6$, we will get $|\tilde{a} - \hat{a}| \leq \xi^{(4)} - \xi^{(3)}$.
- If we switch $\tilde{\xi}$ with $\xi_3$ (which is $\hat{a}$), we will get $|\tilde{a} - \hat{a}| \leq \max\{\xi^{(4)} - \xi^{(3)}, \xi^{(5)} - \xi^{(4)}\}$.

Thus, regardless of which $\xi_i$ is replaced, $|\tilde{a} - \hat{a}|$ is controlled by the maximum spacing between adjacent order statistics near $\hat{a}$.

For order statistics $\xi^{(\ell)}$ from a sample of size $N$ with CDF $F$, the expected difference between consecutive order statistics satisfies:

$$\begin{aligned}
\mathbb{E}\left[\xi^{(\ell+1)}-\xi^{(\ell)}\right] &= \int_0^1 \mathbb{E}\left[\mathbb{1}(\xi^{(\ell+1)}>\xi)-\mathbb{1}(\xi^{(\ell)}>\xi)\right]d\xi \\
&= \int_0^1 \Pr\left[\xi^{(\ell)}\le \xi < \xi^{(\ell+1)}\right]d\xi \\
&= \int_0^1 \binom{N}{\ell} F(\xi)^{\ell}(1-F(\xi))^{N-\ell}d\xi \\
&\le \binom{N}{\ell}\sup_{\xi} F(\xi)^{\ell}(1-F(\xi))^{N-\ell} \\
&= \binom{N}{\ell}\sup_{F\in[0,1]} F^{\ell}(1-F)^{N-\ell}
\end{aligned}$$

which is maximized when $F=\frac{\ell}{N}$:

$$\begin{aligned}
\mathbb{E}\left[\xi^{(\ell+1)}-\xi^{(\ell)}\right] &\le \binom{N}{\ell}\sup_{F\in[0,1]} F^{\ell}(1-F)^{N-\ell} \\
&\le \binom{N}{\ell}\left(\frac{\ell}{N}\right)^{\ell}\left(1-\frac{\ell}{N}\right)^{N-\ell} \\
&\le \binom{N}{N/2}\frac{1}{2^N} \\
&= \Pr\left[\mathrm{Bin}(N,\frac{1}{2})=\frac{N}{2}\right] \\
&= \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

Therefore, we can show that the regret vanishes at rate of $\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$:

$$\mathbb{E}[L(A)-L(a^*)] \;\le\; \mathbb{E}\left[\max\left\{\xi^{(\lfloor qN\rfloor+2)}-\xi^{(\lfloor qN\rfloor+1)},\xi^{(\lfloor qN\rfloor+1)}-\xi^{(\lfloor qN\rfloor)}\right\}\right] \;=\; \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

**Bibliographical notes**. Data-driven Newsvendor and the first proof presented here originate from Levi et al. (2007). The second and fourth proofs presented are original, leveraging statistical learning theory (see Shalev-Shwartz and Ben-David, 2014). The third and fifth proofs also leverage learning theory and come from Xie et al. (2024) and Xie (2024) respectively.

# References

R. Levi, R. O. Roundy, and D. B. Shmoys. Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, 32(4):821–839, 2007.

S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms.* Cambridge university press, 2014.

Y. Xie. Private communication. 2024.

Y. Xie, W. Ma, and L. Xin. Vc theory for inventory policies. *arXiv preprint arXiv:2404.11509*, 2024.