Single Product Dynamic Pricing

- Finite Selling Horizon
- x_0 initial inventory
- dynamically post price π_t at each time $t \in [0,T)$

Inventory Process X_t

Sales Process $N_t \triangleq X_0 - X_t$

- customers arrive according to a Poisson process with rate λ
- i.i.d. valuation with c.d.f. $F(\cdot)$ and p.d.f. $f(\cdot)$
  (U)

$$\mathbb{P}(U \geq p) = 1 - F(p) = \bar{F}(p) \longrightarrow \text{purchasing prob.}$$

Key Assumption

$$v - \bar{F}(v)/f(v) \nearrow \quad v \in \mathbb{R}_+ \text{ w/ a root } v^*.$$

Note that most commonly used functions (uniform, exp, normal, lognormal) ...
satisfies this assumption.

Define the quantile function $q = \bar{F}(p)$

Then the expected revenue rate $R(q) = p\bar{F}(p) = \bar{F}^{-1}(q) \cdot q$

is a concave function in q by the Key Assumption.

$$\frac{dR(q)}{dq} = \frac{dR(q)}{dp} \cdot \frac{dp}{dq} = (\bar{F}(p) - pf(p)) \frac{1}{-f(p)}$$

$$= p - \frac{\bar{F}(p)}{f(p)} \quad \nearrow p \quad \downarrow q$$

Seller's OPT. problem

$$\max_{\pi \in \Pi} \; E\left[\int_{t=0}^{T} \pi_t \, dN_t\right]$$

$$\text{s.t.} \quad \int_{t=0}^{T} dN_t \leq x_0 \qquad \text{a.s.}$$

Let π^* be the opt. policy.

Let the value function be $J^*(x_0, T)$ — x_0 → inventory, T → remaining time

The Bellman eq. says

$$J^*(x, t+\delta) = \sup_{p \in \mathbb{R}_+} \left\{ \lambda \bar{F}(p)\, \delta\, (p + J^*(x-1, t)) + (1 - \lambda \bar{F}(p)\delta) J^*(x,t) + o(\delta) \right\}$$

w/ boundary conditions $J^*(\cdot, 0) = J^*(0, \cdot) = 0$

Let $\delta \to 0$, HJB eq: $\frac{\partial J^*(x,t)}{\partial t} = \sup_{p \in \mathbb{R}_+} \lambda \bar{F}(p)(p + J^*(x-1,t) - J^*(x,t))$

Define

$$\Delta J^*(x,t) = J^*(x,t) - J^*(x-1,t).$$

The opt. policy π^* satisfies the following condition:

$$\pi^*(x,t) - \frac{\bar{F}(\pi^*(x,t))}{f(\pi^*(x,t))} = \Delta J^*(x,t)$$

Results (structural):

- $\forall t$, $\Delta J^*(x,t) \downarrow x$,
- $\forall x$, $\Delta J^*(x,t) \uparrow t$
- $\pi^*(x,t) \geq V^*$ → root of Key Assumption
- $\pi^*(x,t) \downarrow x$
- $\pi^*(x,t) \uparrow t$.

Key Method in RM (fluid approx.)

$\bar{J}(X_0, T)$ by solving

$$\max_{\{\pi_t\}} \int_{t=0}^{T} \lambda \pi_t \bar{F}(\pi_t)\,dt$$
$$\text{s.t. } \int_{t=0}^{T} \lambda \bar{F}(\pi_t)\,dt \le x_0$$

a.s. replaced by expectation.

Key Lemma $\quad J^*(X_0, T) \le \bar{J}(X_0, T)$

Proof: Let

$$J^*(X_0, T, u) = \max_{\pi}\left(E\left[\int_{t=0}^{T} \pi_t\,dN_t\right] - u\left(\int_0^T dN_t - X_0\right)\right)$$

where u is the dual var. for the constr.

$$J^*(X_0, T) \le J^*(X_0, T, u) \quad \forall u \ge 0$$

Lagrangian Relaxation

$$J^*(X_0, T, u) \le \bar{J}(X_0, T, u) \quad \text{by Jensen's eq.}$$

$$\inf_{u \ge 0} \bar{J}(X_0, T, u) = \bar{J}(X_0, T) \quad \text{by zero duality gap.}$$

Note that Jensen's eq. & 0 duality gap requires to see the opt. problem in terms of q_t.

Also, Properties of Point Process give

$$E\left[\int_0^T \pi_t\,dN_t\right] = \int_0^T \lambda \pi_t \bar{F}(\pi_t)\,dt$$
$$E\left[\int_0^T dN_t\right] = \int_0^T \lambda \bar{F}(\pi_t)\,dt.$$

Fixed Price Policy

constr-free price

$$\bar{\pi}_t = \begin{cases} v^*, & \text{if } \lambda T \bar{F}(v^*) \le x_0 \\ \bar{F}^{-1}(x_0/\lambda T), & \text{o/w} \end{cases}$$

constr-binding price.

Theorem (Performance of Fixed Price Policy)

$$R(x_0,T) \geq \frac{J^{\bar{\pi}}(x_0,T)}{J^*(x_0,T)} \geq \frac{J^{\bar{\pi}}(x_0,T)}{\bar{J}(x_0,T)} \geq 1 - \frac{1}{2\sqrt{\min\{\lambda T \bar{F}(v^*), x_0\}}}$$

Implication: $R(kx_0, k\lambda, T) \geq 1 - O(1/\sqrt{k})$

So fixed pricy policy $\bar{\pi}_t$ is asymptotically optimal!

Linear Rate Control (re-optimize at each s)

$$\max_{\{\pi_t\}} \int_{s=t}^{T} \lambda \pi_t \bar{F}(\pi_t)\,dt$$

$$\text{s.t.} \int_{s=t}^{T} \lambda \bar{F}(\pi_t)\,dt \leq \frac{x_0 - \text{inventory consumed}}{T-s}$$

remaining inventory

remaining time

Self-adjusting policy

$$\lambda_s = \lambda \bar{F}(\pi_s) = \left[\bar{\lambda}^* - \sum_{i=1}^{s-1} \left(\Delta_i / T-i\right)\right]^+$$

where $\Delta_i = D_i - \lambda_i$

realized demand

expected demand

Theorem (Jasin 2014)

$$R(kx_0, k\lambda, T) \geq 1 - O\left(\frac{\log k}{k}\right)$$

Analysis of Fixed Pricy Policy

Denote by $\hat{N}$ the total number of customers who arrive during the selling season and are with valuations no less than $\bar{\pi}$. Hence, $\hat{N} \sim \text{Poisson}(\lambda T \bar{F}(\bar{\pi}) \triangleq d)$

$$E[\hat{N}] = \text{Var}[\hat{N}] = d = \lambda T \bar{F}(\bar{\pi}).$$

we then have

$$\begin{aligned}
J^{\bar{\pi}}(X_0, T) &= \bar{\pi}\, E[\min(X_0, \hat{N})] \\
&= \bar{\pi}\, E[\hat{N} - (\hat{N} - X_0)^+] \\
&= \bar{J}(X_0, T) - \bar{\pi} \cdot E[(\hat{N} - X_0)^+] \\
&\geq \bar{J}(X_0, T) - \bar{\pi}\, \frac{\sqrt{d + (X_0 - d)^2} - (X_0 - d)}{2} \\
&\geq \bar{J}(X_0, T) - \frac{\bar{\pi}\sqrt{d}}{2} \\
&= \bar{J}(X_0, T)\left(1 - \frac{1}{2\sqrt{d}}\right) \\
&= \bar{J}(X_0, T)\left(1 - \frac{1}{2\sqrt{\min(\lambda T \bar{F}(v^*), X_0)}}\right)
\end{aligned}$$

Key Inequality:

A r.v. X has mean μ and stdev σ

$$E[(X - a)^+] \leq \frac{\sqrt{\sigma^2 + (a - \mu)^2} - (a - \mu)}{2}$$

Revenue Management : Learn First Then Optimize

Consider a discretized version of Gallego and Van Ryzin '94 wherein the seller sells at time $1, 2, \cdots, T$

The seller does not know λ and $\bar{F}(\cdot)$

- Price is supported on $[0, \bar{p}]$
- Denote $d(p) = \lambda \bar{F}(p)$
- Denote by $N(t)$ the cumulative sales by the end of time t

Key Assumptions

- The virtual value function

$$v - \frac{\bar{F}(v)}{f(v)} \nearrow v$$

- Lipschitz condition $\exists$ constant C such that

$$|p d(p) - p' d(p')| \leq C |p - p'|$$

Learn-first-then-optimize Algorithm

- (learning stage) Implement $K-1$ prices

$\bar{p}/K, \; 2\bar{p}/K, \; \cdots, \; (K-1)\bar{p}/K$ at equal length intervals

Total learning periods $1, 2, \cdots, S$

$$\Pi_t^{LO} = \frac{k}{K} \bar{p}, \text{ if } t \in \left\{ \frac{k-1}{K-1} \cdot S + 1, \cdots, \frac{k}{K-1} \cdot S \right\}, \; k = 1, 2, \cdots, K-1$$

• (estimation) At the end of period s,

estimate the demand rate at price $k\bar{p}/K$ as

$$\hat{d}(k/K\,\bar{p}) = \frac{N_{ks/K-1} - N_{(k-1)s/K-1}}{s/K-1}$$

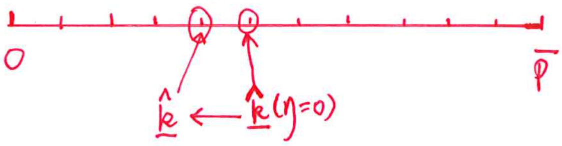
→ cumulative sales

→ length of interval

• (optimization stage) Define

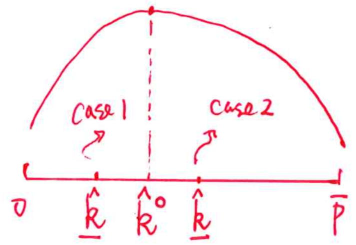
$$\underline{\hat{k}} = \min\left\{k : \hat{d}(k/K\,\bar{p}) \le \frac{(1+\eta)x_0}{T}\right\}$$

"lower" constr. binding price due to η buffer.



Also define

$$\hat{k}^* = \underset{k\in\{\underline{\hat{k}},\cdots,K-1\}}{\arg\max} \ (k/K\,\bar{p})\,\hat{d}(k/K\,\bar{p})$$



If $\hat{k}^o \ge \underline{\hat{k}}$, choose $\hat{k}^o$

o/w, choose $\underline{\hat{k}}$

so $\hat{k}^* = \max(\underline{\hat{k}}, \hat{k}^o)$

During $\{s+1,\cdots,T\}$, the seller posts the price

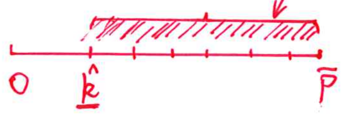
$$\pi_t^{LO} = \hat{p}^* = \hat{k}^*/K\,\bar{p}$$

stops selling when inventory is out.

Regret Analysis

Define

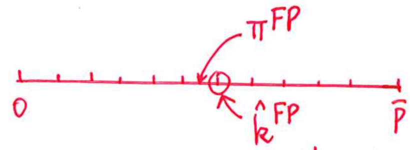
$$g_1 = \mathbb{1}\left\{\pi^{FP} \ge \frac{\underline{\hat{k}}-1}{K}\bar{p}\right\}$$



$$g_2 = \mathbb{1}\left\{\hat{d}(\hat{p}^*) - d(\hat{p}^*) \ge -\eta\, x_0/T\right\}$$

→ Estimated demand and true demand suff. close.

$$\hat{k}^{FP} = \min\{k : \pi^{FP} < k/K\,\bar{p}\}$$



closest to π^{FP} on the grid (on the right)

$\hat{N}_t$ = cumulative sales up to time t assuming no inventory constraint.

$$J^{LO}(X_0,T) \geq \underbrace{\sum_{k=1}^{K-1} \frac{k}{K}\bar{p}\, d\left(\frac{k}{K}\bar{p}\right)\frac{S}{K-1}}_{\text{exploration gain}} + \underbrace{E[\hat{p}^* d(\hat{p}^*)](T-S)}_{\text{exploitation gain}} - \bar{p}\underbrace{E[(\hat{N}_T - X_0)^+]}_{\text{"oversold" quantity}}$$

$$\geq E[\hat{p}^* d(\hat{p}^*)](T-S) - \bar{p}\,E[(\hat{N}_T - X_0)^+]$$

$$= E[\hat{p}^* d(\hat{p}^*)](T-S) - E[\hat{p}^*(\hat{d}(\hat{p}^*) - d(\hat{p}^*))](T-S) - \bar{p}E[(\hat{N}_T - X_0)^+]$$

$$\geq E[\hat{p}^*\hat{d}(\hat{p}^*) \mid g_1]\,\mathbb{P}(g_1)(T-S) - E[\hat{p}^*(\hat{d}(\hat{p}^*) - d(\hat{p}^*))](T-S) - \bar{p}E[(\hat{N}_T - X_0)^+]$$

$$\geq \underbrace{E\left[\frac{\hat{k}^{FP}}{K}\bar{p}\,\hat{d}\left(\frac{\hat{k}^{FP}}{K}\bar{p}\right) \Big| g_1\right]\mathbb{P}(g_1)(T-S)}_{A_1} - \underbrace{\bar{p}E[(\hat{d}(\hat{p}^*) - d(\hat{p}^*))^+](T-S)}_{A_2} - \underbrace{\bar{p}E[(\hat{N}_T - X_0)^+]}_{A_3}$$

(1) Bound A_1

Lipschitz assumption on revenue rate.

$$A_1 \geq \left(\pi^{FP} d(\pi^{FP}) - \frac{C\bar{p}}{K}\right)\mathbb{P}(g_1)(T-S)$$

$$= \pi^{FP} d(\pi^{FP})\,\mathbb{P}(g_1)(T-S) - \frac{C\bar{p}}{K}\mathbb{P}(g_1)(T-S)$$

$$\geq \pi^{FP} d(\pi^{FP})\,T\left(1 - \mathbb{P}(g_1^c) - S/T\right) - \underbrace{\frac{C\bar{p}}{K}\cdot T}_{\text{loss due to discretization}}$$

(2) Bound A_2

$$A_2 \leq E[(\hat{d}(\hat{p}^*) - d(\hat{p}^*))^+]\,T$$

$$\leq \frac{1}{2}\left(\sqrt{\frac{K-1}{S}}\right)T$$

$$\mathrm{Var}(\hat{d}(\hat{p}^*)) = \mathrm{Var}\left(\frac{\frac{S}{K-1}\ \text{Bernoulli sum}}{S/K-1}\right) \leq \frac{K-1}{S}$$

Recall $E[(X-a)^+] \leq \dfrac{\sqrt{\sigma^2 + (a-u)^2} - (a-u)}{2}$

Conditional on g_1,

$$\hat{k}^{FP} = \underset{k \in \{\underline{\hat{k}}, \cdots, K-1\}}{\operatorname{argmax}} \left(\frac{k}{K}\bar{p}\right) d\left(\frac{k}{K}\bar{p}\right)$$

(3) Bound A_3

$$A_3 = E[(\tilde{N}_S + \tilde{N}_T - \tilde{N}_S - x_0)^+]$$

$$\leq E[\tilde{N}_S] + E[(\tilde{N}_T - \tilde{N}_S - x_0)^+]$$

$$\leq S + E[(\tilde{N}_T - \tilde{N}_S - x_0)^+]$$

$$\leq S + E[(\text{Binomial}(T-S, d(\hat{p}^*)) - x_0)^+]$$

$$= S + E[(\text{Binomial}(T-S, d(\hat{p}^*)) - x_0)^+ \mid G_2]\, \mathbb{P}(G_2) + E[(\text{Binomial}(T-S, d(\hat{p}^*)) - x_0)^+ \mid G_2^c]\, \mathbb{P}(G_2^c)$$

$$\leq S + E[(\text{Binomial}(T-S, \min\{\hat{d}(\hat{p}^*) + \frac{\eta x_0}{T}, 1\}) - x_0)^+ \mid G_2] + (T-S)\mathbb{P}(G_2^c)$$

$$\leq S + E[(\text{Binomial}(T-S, \min\{\frac{(1+2\eta)x_0}{T}, 1\}) - x_0)^+] + (T-S)\mathbb{P}(G_2^c)$$

$$\leq S + E[(\text{Binomial}(T, \min\{\frac{(1+2\eta)x_0}{T}, 1\}) - x_0)^+] + (T-S)\mathbb{P}(G_2^c)$$

$$\leq S + \frac{1}{2}\left(\sqrt{(1+2\eta)x_0 + 4\eta^2 x_0^2} + 2\eta x_0\right) + (T-S)\mathbb{P}(G_2^c)$$

$$\leq S + \frac{1}{2}\left(\sqrt{(1+2\eta)x_0 + 4\eta^2 x_0^2} + 2\eta x_0\right) + (T-S)\exp\left(-\left(\frac{2\eta^2 x_0^2}{T^2}\right)\left(\frac{S}{K-1}\right)\right)$$

Avg. of $S/(K-1)$ Bernoulli R.V.'s

$$\mathbb{P}(G_2^c) = \mathbb{P}(\hat{d}(\hat{p}^*) - d(\hat{p}^*) \geq -\frac{\eta x_0}{T})$$

$$\leq e^{\left(-2\left(\frac{S}{K-1}\right)\left(\frac{\eta x_0}{T}\right)^2\right)}$$

It remains to bound $\mathbb{P}(\mathcal{G}_1^c)$

$$\mathbb{P}(\mathcal{G}_1^c) \le \mathbb{P}\left(d\left(\frac{\hat{k}-1}{K}\bar{p}\right) \le x_0/T\right) \quad \text{By def. of } \hat{k}$$

$$\le \mathbb{P}\left(\hat{d}\left(\frac{\hat{k}-1}{K}\bar{p}\right) - d\left(\frac{\hat{k}-1}{K}\bar{p}\right) \ge \frac{\eta x_0}{T}\right) \quad \text{Hoeffding's ineq}$$

$$\le \exp\left(-2\left(\frac{\eta^2 x_0^2}{T^2}\right)\left(\frac{k-1}{s}\right)\right)$$

$$\text{Regret} \le \pi^{FP} d(\pi^{FP}) \exp\left(-\frac{2\eta^2 x_0^2 s}{T^2(k-1)}\right) + \pi^{FP} d(\pi^{FP}) s$$

$$+ \frac{C\bar{p}}{K} T + \frac{\bar{p}}{2}\sqrt{\frac{k-1}{s}}\, T + \bar{p}s$$

$$+ \frac{\bar{p}}{2}\left(\sqrt{(1+2\eta)x_0 + 4\eta^2 x_0^2} + 2\eta x_0\right) + \bar{p}(T-s)\exp\left(-\frac{2\eta^2 x_0^2 s}{T^2(k-1)}\right)$$

$$\le \left(\pi^{FP} d(\pi^{FP}) + \bar{p}\right) T \exp\left(-\frac{2\eta^2 x_0^2 s}{T^2(k-1)}\right) + \left(\pi^{FP} d(\pi^{FP}) + \bar{p}\right) s$$

$$+ \frac{C\bar{p}}{K} \cdot T + \frac{\bar{p}}{2}\sqrt{\frac{k-1}{s}}\, T + \frac{\bar{p}}{2}\left(\sqrt{(1+2\eta)x_0 + 4\eta^2 x_0^2} + 2\eta x_0\right)$$

Consider a seq. of problems indexed by n.

In the nth problem, $x_0^{(n)} = n x_0$.

$T^{(n)} = nT$.

set $\quad s^{(n)} = s n^{3/4}, \quad K^{(n)} = K n^{1/4}, \quad \eta^{(n)} = (\log n)^{1/2} n^{-1}$

$$\text{Regret}^{(n)} \le O(1) + O(n^{3/4}) + O(n^{3/4}) + O(n^{3/4}) + O((\log n)^{1/2} n^{3/4})$$

$$= O\left((\log n)^{1/2} n^{3/4}\right)$$

Revenue Management : UCB

Consider a discretized version of Gallego and van Ryzin '94 wherein the seller sells at time $1, 2, \cdots, T$ and the prices $\in \{p_1, p_2, \cdots, p_K\}$ upper bounded by $\bar{p}$

WLOG, let $\bar{p} = 1$.

At price p_k, $\mathbb{P}(\text{a sale occurs}) = \theta_k$

- The seller does not know $\{\theta_1, \theta_2, \cdots, \theta_K\}$.

Denote by N_t the cumulative sales by the end of time t.

Fluid Approximation with full Information

$$\max_{\{q_1, q_2, \cdots, q_K\}} \quad g_\theta(q) \triangleq \sum_{k=1}^{K} p_k \theta_k q_k$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \theta_k q_k \leq x_0 / T$$

$$\sum_{k=1}^{K} q_k \leq 1$$

$$q_k \geq 0 \quad \forall k.$$

The optimal solution is denoted as q^*.

UCB Algorithm

For each price p_k,

$$A_{t-1}(k) = \#\text{ price } p_k \text{ has been posted up to time } t-1$$
$$= \sum_{s=1}^{t-1} \mathbb{1}(\pi_s = p_k)$$

$$N_{t-1}(k) = \#\text{ sales under price } p_k \text{ up to time } t-1$$
$$= \sum_{s=1}^{t-1} (N_s - N_{s-1})\, \mathbb{1}(\pi_s = p_k)$$

Define the upper confidence sales rate as

$$\bar{\theta}_t(k) = \min\left\{ \frac{N_{t-1}(k)}{A_{t-1}(k)} + \beta \sqrt{\frac{\log T}{A_{t-1}(k)}},\ 1 \right\}, \quad \beta > 0$$

the lower confidence sales rate as

$$\underline{\theta}_t(k) = \max\left\{ \frac{N_{t-1}(k)}{A_{t-1}(k)} - \beta \sqrt{\frac{\log T}{A_{t-1}(k)}},\ 0 \right\}$$

Consider the following upper confidence optimization problem:

$$\max_{\{q_1, q_2, \cdots q_K\}} \quad U_t(q) \triangleq \sum_{k=1}^{K} p_k \bar{\theta}_t(k) q_k$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \underline{\theta}_t(k) q_k \le x_0 / T$$

$$\sum_{k=1}^{K} q_k \le 1$$

$$q_k \ge 0 \quad \forall k.$$

The optimal solution is denoted by q_t^{UCB}

Consider the following algorithm:

- At time $t=1,2,\cdots,K$, the seller posts price p_k.
- At time $t=K+1,\cdots,T$, the seller posts price p_k w.p. $q_t^{UCB}(k)$

The seller stops selling when inventory is out.

Regret Analysis

(1) $T \leq K$. Regret $\leq g_\theta(q^*)T \leq T$.

(2) $T > K$. We allow the seller to keep on applying UCB even when the inventory is out.

Denote by $\tilde{N}_t$ the cumulative sales up to time t without inventory constraint.

Define the lower confidence revenue function

$$L_t(q) \triangleq \sum_{k=1}^{K} p_k \underline{\theta}_k(k) q_k$$

$$\begin{aligned}
\text{Regret} &= g_\theta(q^*)T - E\left[\sum_{t=1}^{T} g_\theta(q_t^{UCB})\right] + E\left[(\tilde{N}_T - x_0)^+\right] \\
&= \sum_{t=1}^{T} E\left[g_\theta(q^*) - g_\theta(q_t^{UCB})\right] + E\left[(\tilde{N}_T - x_0)^+\right] \\
&= \sum_{t=1}^{K} E\left[g_\theta(q^*) - g_\theta(q_t^{UCB})\right] + \sum_{t=K+1}^{T} E\left[g_\theta(q^*) - g_\theta(q_t^{UCB})\right] \\
&\quad + E\left[(\tilde{N}_T - x_0)^+\right] \\
&\leq K + \sum_{t=K+1}^{T} E\left[g_\theta(q^*) - g_\theta(q_t^{UCB})\right] + E\left[(\tilde{N}_T - x_0)^+\right]
\end{aligned}$$

$$= K + \sum_{t=K+1}^{T} E[g_\theta(q^*) - U_t(q_t^{UCB}) + U_t(q_t^{UCB}) - L_t(q_t^{UCB}) + L_t(q_t^{UCB}) - g_\theta(q_t^{UCB})] + E[(\hat{N}_T - X_0)^+]$$

$$= K + \underbrace{\sum_{t=K+1}^{T} E[g_\theta(q^*) - U_t(q_t^{UCB})]}_{A} + \underbrace{\sum_{t=K+1}^{T} E[U_t(q_t^{UCB}) - L_t(q_t^{UCB})]}_{B} + \underbrace{\sum_{t=K+1}^{T} E[L_t(q_t^{UCB}) - g_\theta(q_t^{UCB})]}_{C} + \underbrace{E[(\hat{N}_T - X_0)^+]}_{D}$$

Bound (A)

$$A = \underbrace{\sum_{t=K+1}^{T} E[g_\theta(q^*) - U_t(q^*)]}_{A_1} + \underbrace{\sum_{K+1}^{T} E[U_t(q^*) - U_t(q_t^{UCB})]}_{A_2}$$

$$A_1 = \sum_{t=K+1}^{T} E[(g_\theta(q^*) - U_t(q^*)) \mathbb{1}(g_\theta(q^*) > U_t(q^*))] + \sum_{t=K+1}^{T} E[(g_\theta(q^*) - U_t(q^*)) \mathbb{1}(g_\theta(q^*) \le U_t(q^*))]$$

$$\le \sum_{t=K+1}^{T} E[(g_\theta(q^*) - U_t(q^*)) \mathbb{1}(g_\theta(q^*) > U_t(q^*))]$$

$$\le \sum_{t=K+1}^{T} P(g_\theta(q^*) > U_t(q^*))$$

Hoeffding's ineq.

$$\le \sum_{t=K+1}^{T} e^{-2\beta^2 \log T}$$

$$= \sum_{t=K+1}^{T} \frac{1}{T^{2\beta^2}} \le T^{1-2\beta}$$

Define

$$\mathcal{G}_t \triangleq \mathbb{1}\left(\sum_{k=1}^{K} \underline{\theta}_t(k) q^* \leq x_0/T\right)$$

$$A_2 = \sum_{t=K+1}^{T} E[U_t(q^*) - U_t(q_t^{UCB}) \mid \mathcal{G}_t]\, \mathbb{P}(\mathcal{G}_t) +$$

$$\sum_{t=K+1}^{T} E[U_t(q^*) - U_t(q_t^{UCB}) \mid \mathcal{G}_t^c]\, \mathbb{P}(\mathcal{G}_t^c)$$

$$\leq \sum_{t=K+1}^{T} \mathbb{P}(\mathcal{G}_t^c)$$

$$\leq \sum_{t=K+1}^{T} \mathbb{P}\left(\sum_{k=1}^{K} \underline{\theta}_t(k) q^* > \sum_{k=1}^{K} \underline{\theta}_k q^*\right)$$

$$\leq \sum_{t=K+1}^{T} e^{-2\beta^2 \log T} = \sum_{t=K+1}^{T} 1/T^{2\beta^2} \leq T^{1-2\beta}$$

Bound (c)

$$C = \sum_{t=K+1}^{T} E[(L_t(q_t^{UCB}) - g_\theta(q_t^{UCB}))\, \mathbb{1}(L_t(q_t^{UCB}) > g_\theta(q_t^{UCB}))]$$

$$+ \sum_{t=K+1}^{T} E[(L_t(q_t^{UCB}) - g_\theta(q_t^{UCB}))\, \mathbb{1}(L_t(q_t^{UCB}) \leq g_\theta(q_t^{UCB}))]$$

$$\leq \sum_{t=K+1}^{T} E[(L_t(q_t^{UCB}) - g_\theta(q_t^{UCB}))\, \mathbb{1}(L_t(q_t^{UCB}) > g_\theta(q_t^{UCB}))]$$

$$\leq \sum_{t=K+1}^{T} \mathbb{P}(L_t(q_t^{UCB}) > g_\theta(q_t^{UCB}))$$

$$\leq \sum_{t=K+1}^{T} e^{-2\beta^2 \log T} = \sum_{t=K+1}^{T} 1/T^{2\beta^2} \leq T^{1-2\beta^2}$$

Bound (B)

Define $\mathcal{T}_k = \{ t : t \in \{K+1, \cdots, T\},\ \pi_t^{UCB} = p_k \}$

Then

$$
\begin{aligned}
B &\leq E\left[\sum_{k\in\{1,\cdots,K\}} \sum_{t\in\mathcal{T}_k} 2\beta \sqrt{\frac{\log T}{A_{t-1}(k)}} \right] \\
&= 2\beta\sqrt{\log T}\ E\left[\sum_{k\in\{1,2,\cdots,K\}} \sum_{t\in\mathcal{T}_k} \frac{1}{\sqrt{A_{t-1}(k)}} \right] \\
&= 2\beta\sqrt{\log T}\ E\left[\sum_{k\in\{1,\cdots,K\}} \sum_{n=1}^{|\mathcal{T}_k|} \frac{1}{\sqrt{n}} \right] \\
&\leq 2\beta\sqrt{\log T}\ E\left[\sum_{k\in\{1,\cdots,K\}} \int_{n=0}^{|\mathcal{T}_k|} \frac{1}{\sqrt{n}}\, dn \right] \\
&= 2\beta\sqrt{\log T}\ E\left[\sum_{k\in\{1,\cdots,K\}} 2\sqrt{|\mathcal{T}_k|} \right] \\
&\leq 4\beta\sqrt{K(T-K)\log T} \leq 4\beta\sqrt{KT\log T}
\end{aligned}
$$

Bound (D)

$$D \le E\Big[\Big(K + \sum_{t=K+1}^{T} \text{Bernoulli}\Big(\sum_{k=1}^{K} \theta_k q_{t,k}^{UCB}\Big) - x_0\Big)^+\Big]$$

$$\le E\Big[\Big(K + \sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} (\underline{\theta}_t(k) + (\theta_k - \underline{\theta}_t(k))^+ q_{t,k}^{UCB}, 1\Big\}\Big) - x_0\Big)^+\Big]$$

$$\le K + E\Big[\Big(\sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} (\underline{\theta}_t(k) + (\theta_k - \underline{\theta}_t(k))^+ q_{t,k}^{UCB}, 1\Big\}\Big) - x_0 \frac{T-K}{T}\Big)^+\Big]$$

$$\le K + E\Big[\Big(\sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} \underline{\theta}_t(k) q_{t,k}, 1\Big\}\Big)$$

$$+ \sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} ((\theta_k - \underline{\theta}_t(k))^+ q_{t,k}^{UCB}, 1\Big\}\Big) - x_0 \frac{T-K}{T}\Big)^+\Big]$$

$$\le K + E\Big[\Big(\sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{x_0/T, 1\Big\}\Big)$$

$$+ \sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} ((\theta_k - \underline{\theta}_t(k))^+, 1\Big\}\Big) - x_0 \frac{T-K}{T}\Big)^+\Big]$$

$$\le K + E\Big[\Big(\sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{x_0/T, 1\Big\}\Big) - x_0 \frac{T-K}{T}\Big)^+\Big]$$

$$+ E\Big[\sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} ((\theta_k - \underline{\theta}_t(k))^+, 1\Big\}\Big)\Big]$$

$$= K + E\Big[\Big(\text{Binomial}\Big(T-K, \min\Big\{x_0/T, 1\Big\}\Big) - x_0 \frac{T-K}{T}\Big)^+\Big]$$

$$+ E\Big[\sum_{t=K+1}^{T} \text{Bernoulli}\Big(\min\Big\{\sum_{k=1}^{K} (\theta_k - \underline{\theta}_t(k))^+, 1\Big\}\Big)\Big]$$

$$\le K + \frac{\sqrt{T-K}}{2} + \sum_{t=K+1}^{T} E\Big[\min\Big\{\sum_{k=1}^{K} (\theta_k - \underline{\theta}_t(k))^+, 1\Big\}\Big]$$

$$\le K + \sqrt{T}/2 + \sum_{t=K+1}^{T} \sum_{k=1}^{K} E[(\theta_k - \underline{\theta}_t(k))^+]$$

$$= K + \sqrt{T/2} + \sum_{t=K+1}^{T} \sum_{k=1}^{K} E\Big[E\big[(\theta_k - \underline{\theta}_t(k))^+ \,\big|\, A_{t-1}(k)\big]\Big]$$

$$\le K + \sqrt{T/2} + \frac{1}{2} \sum_{t=K+1}^{T} \sum_{k=1}^{K} E\left[\sqrt{\frac{\theta_k}{A_{t-1}(k)} + \beta^2 \frac{\log T}{A_{t-1}(k)}} + \beta \sqrt{\frac{\log T}{A_{t-1}(k)}} \right]$$

$$\le K + \sqrt{T/2} + \frac{1 + 2\beta\sqrt{\log T}}{2} \sum_{t=K+1}^{T} \sum_{k=1}^{K} E\left[\frac{1}{\sqrt{A_{t-1}(k)}} \right]$$

$$\le K + \sqrt{T/2} + \sqrt{KT} + 2\beta\sqrt{KT \log T}$$

Thus,

$$\text{Regret} \le 2K + 3T^{1-2\beta^2} + 6\beta\sqrt{KT\log T} + \sqrt{T/2} + \sqrt{KT}$$

By taking $\beta = 1/2$, we have

$$\text{Regret} \le 2K + 3.5\sqrt{T} + 3\sqrt{KT\log T} + \sqrt{KT}.$$

Recall

$$E[(X-a)^+] \le \frac{\sqrt{\sigma^2 + (a-\mu)^2} - (a-\mu)}{2}$$