Dynamic pricing and learning

Arnoud den Boer

University of Twente

Lunteren, January 13, 2015
Learning in sequential decision problems

OPTIMIZATION

Determine optimal decision
Learning in sequential decision problems

**OPTIMIZATION**
Determine optimal decision

**STATISTICS**
Estimate unknown parameters
Learning in sequential decision problems

OPTIMIZATION
Determine optimal decision

Estimate unknown parameters

STATISTICS DATA
Learning in sequential decision problems

**OPTIMIZATION**

Determine optimal decision

- Estimate unknown parameters
- Generate new data

**STATISTICS** → **DATA**
Learning in sequential decision problems

OPTIMIZATION
Determine optimal decision

STATISTICS
Estimate unknown parameters

DATA
Generate new data
Dynamic pricing

- A firm sells a single product in discrete time periods $t = 1, \ldots, T$. 
Dynamic pricing

- A firm sells a single product in discrete time periods \( t = 1, \ldots, T \).
- Each period \( t \): (i) choose selling price \( p_t \);
  (ii) observe demand

\[
d_t = \theta_1 + \theta_2 p_t + \epsilon_t,
\]

where \( \theta = (\theta_1, \theta_2) \) are unknown parameters in known set \( \Theta \),
\( \epsilon_t \) unobservable random disturbance term;
(iii) collect revenue \( p_t d_t \).

Which non-anticipating prices \( p_1, \ldots, p_T \) maximize cumulative expected revenue

\[
\min_{\theta \in \Theta} E \left[ \sum_{t=1}^{T} p_t d_t \right]
\]

Intractable problem
A firm sells a single product in discrete time periods $t = 1, \ldots, T$.

Each period $t$: (i) choose selling price $p_t$;
(ii) observe demand

$$d_t = \theta_1 + \theta_2 p_t + \epsilon_t,$$

where $\theta = (\theta_1, \theta_2)$ are unknown parameters in known set $\Theta$, $\epsilon_t$ unobservable random disturbance term;
(iii) collect revenue $p_t d_t$.

Which non-anticipating prices $p_1, \ldots, p_T$ maximize cumulative expected revenue $\min_{\theta \in \Theta} E \left[ \sum_{t=1}^{T} p_t d_t \right]$?
Dynamic pricing

- A firm sells a single product in discrete time periods \( t = 1, \ldots, T \).
- Each period \( t \): (i) choose selling price \( p_t \); (ii) observe demand
  \[
  d_t = \theta_1 + \theta_2 p_t + \epsilon_t,
  \]
  where \( \theta = (\theta_1, \theta_2) \) are unknown parameters in known set \( \Theta \), \( \epsilon_t \) unobservable random disturbance term; (iii) collect revenue \( p_t d_t \).
- Which non-anticipating prices \( p_1, \ldots, p_T \) maximize cumulative expected revenue \( \min_{\theta \in \Theta} \mathbb{E} \left[ \sum_{t=1}^{T} p_t d_t \right] \)?

Intractable problem
Myopic pricing

An intuitive solution

- Choose arbitrary initial prices $p_1 \neq p_2$.
- For each $t \geq 2$:
  (i) determine LS estimate $\hat{\theta}_t$ of $\theta$, based on available sales data;
  (ii) set

$$p_{t+1} = \arg \max_p p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2}p)$$
Myopic pricing

An intuitive solution

- Choose arbitrary initial prices $p_1 \neq p_2$.
- For each $t \geq 2$:
  (i) determine LS estimate $\hat{\theta}_t$ of $\theta$, based on available sales data;
  (ii) set

$$p_{t+1} = \arg \max_p p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2} p)$$

perceived optimal decision
An intuitive solution

- Choose arbitrary initial prices \( p_1 \neq p_2 \).
- For each \( t \geq 2 \):
  (i) determine LS estimate \( \hat{\theta}_t \) of \( \theta \), based on available sales data;
  (ii) set

\[
p_{t+1} = \arg \max_p \ p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2}p)
\]

perceived optimal decision

- ‘Always choose the perceived optimal action’.
Does $\hat{\theta}_t$ converge to $\theta$ as $t \to \infty$?
Does $\hat{\theta}_t$ converge to $\theta$ as $t \to \infty$?

No

It seems that $\hat{\theta}_t$ always converges, but w.p. zero to the true $\theta$. Open problem.
Convergence

Does $\hat{\theta}_t$ converge to $\theta$ as $t \to \infty$?

No

It seems that $\hat{\theta}_t$ always converges, but w.p. zero to the true $\theta$.
Open problem.

Caused by the prevalence of indeterminate equilibria:
Parameter estimates such that the true expected demand at the myopic optimal price equals the predicted expected demand.
If $\hat{\theta}$ suff. close to $\theta$, then $\arg\max_p p \cdot (\hat{\theta}_1 + \hat{\theta}_2 p) = -\hat{\theta}_1/(2\hat{\theta}_2)$.

Then:

'True' expected demand: $\theta_1 + \theta_2 -\frac{\hat{\theta}_1}{2\hat{\theta}_2}$. \hspace{1cm} (1)

'Predicted' expected demand: $\hat{\theta}_1 + \hat{\theta}_2 -\frac{\hat{\theta}_1}{2\hat{\theta}_2}$. \hspace{1cm} (2)

If (1) equals (2), then $\hat{\theta}$ is an IE.

Model output 'confirms' correctness of the (incorrect) estimates.
Indeterminate equilibria

If $\hat{\theta}$ suff. close to $\theta$, then $\arg\max_{p} p \cdot (\hat{\theta}_1 + \hat{\theta}_2 p) = -\hat{\theta}_1 / (2\hat{\theta}_2)$.

Then:

‘True’ expected demand: $\theta_1 + \theta_2 \frac{-\hat{\theta}_1}{2\hat{\theta}_2}$.  \hspace{1cm} (1)

‘Predicted’ expected demand: $\hat{\theta}_1 + \hat{\theta}_2 \frac{-\hat{\theta}_1}{2\hat{\theta}_2}$. \hspace{1cm} (2)

If (1) equals (2), then $\hat{\theta}$ is an IE.
Model output ‘confirms’ correctness of the (incorrect) estimates.
Indeterminate equilibria: example

Indeterminate Equilibria - Dynamic Pricing

Plot showing the relationship between two variables, $a_1$ and $a_2$, with data points and a trend line.
Back to original problem

Which non-anticipating prices $p_1, \ldots, p_T$ maximize

$$\min_{\theta \in \Theta} \mathbb{E} \left[ \sum_{t=1}^{T} p_t d_t \right],$$

or, equivalently, minimize the Regret$(T)$

$$\max_{\theta \in \Theta} \mathbb{E} \left[ T \cdot \max_p p \cdot (\theta_1 + \theta_2 p) - \sum_{t=1}^{T} p_t d_t \right]$$
Back to original problem

Which non-anticipating prices $p_1, \ldots, p_T$ maximize

$$\min_{\theta \in \Theta} \mathbb{E}\left[ \sum_{t=1}^{T} p_t d_t \right],$$

or, equivalently, minimize the Regret$(T)$

$$\max_{\theta \in \Theta} \mathbb{E}\left[ T \cdot \max_{p} p \cdot (\theta_1 + \theta_2 p) - \sum_{t=1}^{T} p_t d_t \right]$$

- Exact solution intractable
Back to original problem

Which non-anticipating prices $p_1, \ldots, p_T$ maximize

$$\min_{\theta \in \Theta} \mathbb{E}\left[ \sum_{t=1}^{T} p_t d_t \right],$$

or, equivalently, minimize the Regret($T$)

$$\max_{\theta \in \Theta} \mathbb{E}\left[ T \cdot \max_{p} p \cdot (\theta_1 + \theta_2 p) - \sum_{t=1}^{T} p_t d_t \right]$$

- Exact solution intractable
- Myopic pricing not optimal
Which non-anticipating prices $p_1, \ldots, p_T$ maximize

$$\min_{\theta \in \Theta} \mathbb{E}\left[ \sum_{t=1}^{T} p_t d_t \right],$$

or, equivalently, minimize the Regret($T$)

$$\max_{\theta \in \Theta} \mathbb{E}\left[ T \cdot \max_p p \cdot (\theta_1 + \theta_2 p) - \sum_{t=1}^{T} p_t d_t \right]$$

- Exact solution intractable
- Myopic pricing not optimal
- Let’s find asymptotically optimal policies: smallest growth rate of Regret($T$) in $T$. 
Asymptotically optimal policy

Important observation: Variation in controls $\Rightarrow$ better estimates.
Asymptotically optimal policy

Important observation: Variation in controls $\Rightarrow$ better estimates.

$$\| \hat{\theta}_t - \theta \|^2 = O \left( \frac{\log t}{t \text{Var}(p_1, \ldots, p_t)} \right) \text{ a.s.}$$

To ensure convergence of $\hat{\theta}_t$, some amount of experimentation is necessary.
Asymptotically optimal policy

Important observation: Variation in controls $\Rightarrow$ better estimates.

$$\left\| \hat{\theta}_t - \theta \right\|^2 = O \left( \frac{\log t}{t \text{Var}(p_1, \ldots, p_t)} \right) \text{ a.s.}$$

To ensure convergence of $\hat{\theta}_t$, some amount of experimentation is necessary. But, not too much.
Controlled Variance pricing

- Choose arbitrary initial prices $p_1 \neq p_2$.
- For each $t \geq 2$:
  (i) determine LS estimate $\hat{\theta}_t$ of $\theta$, based on available sales data;
  (ii) set
  \[
  p_{t+1} = \arg\max_{p} p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2} p)
  \]
Choose arbitrary initial prices $p_1 \neq p_2$.

For each $t \geq 2$:
  (i) determine LS estimate $\hat{\theta}_t$ of $\theta$, based on available sales data;
  (ii) set
  $$p_{t+1} = \arg \max_p p \cdot (\hat{\theta}_t + \hat{\theta}_t p)$$
  perceived optimal decision

subject to
$$\text{Var}(p_1, \ldots, p_{t+1}) \geq f(t),$$

‘information constraint’ for some increasing $f: \mathbb{N} \to (0, \infty)$.

‘Always choose the perceived optimal action that induces sufficient experimentation’.
‘Controlled Variance pricing’

• Choose arbitrary initial prices \( p_1 \neq p_2 \).
• For each \( t \geq 2 \):
  (i) determine LS estimate \( \hat{\theta}_t \) of \( \theta \), based on available sales data;
  (ii) set

\[
p_{t+1} = \arg \max_p p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2} p) \quad \text{perceived optimal decision}
\]

s.t. \( t \cdot \text{Var}(p_1, \ldots, p_{t+1}) \geq f(t), \quad \text{‘information constraint’} \)
‘Controlled Variance pricing’

- Choose arbitrary initial prices $p_1 \neq p_2$.
- For each $t \geq 2$:
  (i) determine LS estimate $\hat{\theta}_t$ of $\theta$, based on available sales data;
  (ii) set

\[
p_{t+1} = \arg \max_p p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2}p)
\]

perceived optimal decision

\[
\text{s.t. } t \cdot \text{Var}(p_1, \ldots, p_{t+1}) \geq f(t), \quad \text{‘information constraint’}
\]

for some increasing $f : \mathbb{N} \to (0, \infty)$. 
Choose arbitrary initial prices $p_1 \neq p_2$.

For each $t \geq 2$:

(i) determine LS estimate $\hat{\theta}_t$ of $\theta$, based on available sales data;

(ii) set

$$p_{t+1} = \arg \max_p p \cdot (\hat{\theta}_{t1} + \hat{\theta}_{t2}p)$$

perceived optimal decision

s.t. $t \cdot \text{Var}(p_1, \ldots, p_{t+1}) \geq f(t)$, ‘information constraint’

for some increasing $f : \mathbb{N} \rightarrow (0, \infty)$.

‘Always choose the perceived optimal action that induces sufficient experimentation’.
Regret($T$) = $O\left(f(T) + \sum_{t=1}^{T} \frac{\log t}{f(t)}\right)$. 

Optimal $f$ gives Regret($T$) = $O(\sqrt{T \log T})$, and no policy beats $\sqrt{T}$. Thus, you can characterize asymptotically optimal amount of experimentation.
Regret($T$) = $O\left(f(T) + \sum_{t=1}^{T} \frac{\log t}{f(t)}\right)$.

\(f\) balances between exploration and exploitation.
Regret\( (T) = O \left( f(T) + \sum_{t=1}^{T} \frac{\log t}{f(t)} \right) \).

\( f \) balances between exploration and exploitation.

Optimal \( f \) gives Regret\( (T) = O(\sqrt{T \log T}) \), and no policy beats \( \sqrt{T} \).

Thus, you can characterize asymptotically optimal amount of experimentation.
‘Controlled Variance pricing’ - performance

- Regret($T$) = $O\left(f(T) + \sum_{t=1}^{T} \frac{\log t}{f(t)}\right)$.
- $f$ balances between exploration and exploitation.
- Optimal $f$ gives Regret($T$) = $O(\sqrt{T \log T})$, and no policy beats $\sqrt{T}$.

Thus, you can characterize asymptotically optimal amount of experimentation.
Extension: multiple products

\( K \) products: price vector \( \mathbf{p}_t = (p_t(1), \ldots, p_t(K))^\top \),
demand vector \( \mathbf{d}_t = \theta \begin{pmatrix} 1 \\ \mathbf{p}_t \end{pmatrix} + \epsilon \), matrix \( \theta \), noise-vector \( \epsilon \).
Extension: multiple products

\( K \) products: price vector \( \mathbf{p}_t = (p_t(1), \ldots, p_t(K))^\top \),
demand vector \( \mathbf{d}_t = \theta \begin{pmatrix} 1 \\ \mathbf{p}_t \end{pmatrix} + \epsilon \), matrix \( \theta \), noise-vector \( \epsilon \).

Convergence rates of LS-estimator:

\[
\left\| \hat{\theta}_t - \theta \right\|^2 = O \left( \frac{\log t}{\lambda_{\min}(t)} \right) \text{ a.s.,}
\]

where \( \lambda_{\min}(t) \) is the smallest eigenvalue of the information matrix

\[
\sum_{i=1}^{t} \begin{pmatrix} 1 \\ \mathbf{p}_i \\ \mathbf{p}_i \mathbf{p}_i^\top \end{pmatrix}
\]
Extension: multiple products

Same type of policy:

\[ p_{t+1} = \arg \max_p \mathbf{p}^\top \hat{\theta}_t \begin{pmatrix} 1 \\ \mathbf{p} \end{pmatrix} \]
Extension: multiple products

Same type of policy:

\[ p_{t+1} = \arg \max_p p^\top \hat{\theta}_t \left( \begin{array}{c} 1 \\ p \end{array} \right) \text{ perceived optimal decision} \]
Extension: multiple products

Same type of policy:

\[ p_{t+1} = \arg \max_p p^\top \hat{\theta}_t \begin{pmatrix} 1 \\ p \end{pmatrix} \] perceived optimal decision

s.t. \( \lambda_{\text{min}}(t+1) \geq f(t) \), ‘information constraint’
Extension: multiple products

Same type of policy:

\[ p_{t+1} = \arg \max_p p^\top \hat{\theta}_t \left( \begin{array}{c} 1 \\ p \end{array} \right) \]  

perceived optimal decision

s.t. \( \lambda_{\min}(t + 1) \geq f(t) \), \( \text{‘information constraint’} \)

for some increasing \( f : \mathbb{N} \rightarrow (0, \infty) \).
Extension: multiple products

Same type of policy:

\[ p_{t+1} = \arg \max_p p^\top \hat{\theta}_t \left( \begin{array}{c} 1 \\ p \end{array} \right) \quad \text{perceived optimal decision} \]

s.t. \( \lambda_{\min}(t + 1) \geq f(t), \quad \text{‘information constraint’} \)

for some increasing \( f : \mathbb{N} \to (0, \infty) \).

Problem: \( \lambda_{\min}(t + 1) \) is a complicated object.
Extension: multiple products

Same type of policy:

$$p_{t+1} = \arg \max_p p^\top \hat{\theta}_t \left( \begin{array}{c} 1 \\ p \end{array} \right)$$

perceived optimal decision

s.t. $\lambda_{\min}(t+1) \geq f(t)$, ‘information constraint’

for some increasing $f : \mathbb{N} \to (0, \infty)$.

Problem: $\lambda_{\min}(t+1)$ is a complicated object.

Convertible to non-convex but tractable quadratic constraint.
Extension: multiple products

Same type of policy:

\[ p_{t+1} = \arg \max_p p^\top \hat{\theta}_t \begin{pmatrix} 1 \\ p \end{pmatrix} \quad \text{perceived optimal decision} \]

s.t. \( \lambda_{\min}(t + 1) \geq f(t) \), ‘information constraint’

for some increasing \( f : \mathbb{N} \rightarrow (0, \infty) \).

Problem: \( \lambda_{\min}(t + 1) \) is a complicated object.

Convertible to non-convex but tractable quadratic constraint.

\[
\text{Regret}(T) = O \left( f(T) + \sum_{t=1}^{T} \frac{\log t}{f(t)} \right),
\]

optimal \( f \) gives \( \text{Regret}(T) = O(\sqrt{T \log T}) \).
Many more extensions

- Non-linear demand functions, non-parametric estimation
Many more extensions

- Non-linear demand functions, non-parametric estimation
- Time-varying markets (how much data to use for inference?)
Many more extensions

- Non-linear demand functions, non-parametric estimation
- Time-varying markets (how much data to use for inference?)
- Strategic customer behavior (can you detect this from data?)
Many more extensions

- Non-linear demand functions, non-parametric estimation
- Time-varying markets (how much data to use for inference?)
- Strategic customer behavior (can you detect this from data?)
- Competition (repeated games with incomplete information? Mean field games with learning?)
Many more extensions

- Non-linear demand functions, non-parametric estimation
- Time-varying markets (how much data to use for inference?)
- Strategic customer behavior (can you detect this from data?)
- Competition (repeated games with incomplete information? Mean field games with learning?)
Many more extensions

- Non-linear demand functions, non-parametric estimation
- Time-varying markets (how much data to use for inference?)
- Strategic customer behavior (can you detect this from data?)
- Competition (repeated games with incomplete information? Mean field games with learning?)

Very fruitful interaction between theory and practice
Dynamic pricing is an example of discrete-choice problems:

- Decision maker chooses attributes and/or availability of alternatives
More general...

Dynamic pricing is an example of discrete-choice problems:
- Decision maker chooses attributes and/or availability of alternatives
- Customers choose according to probabilistic choice-model

Open PhD Position:
http://www.utwente.nl/ewi/sor/about/staff/boer/
Dynamic pricing is example of **discrete-choice problems:**

- Decision maker chooses attributes and/or availability of alternatives
- Customers choose according to probabilistic choice-model
- Decision maker aims to optimize some objective function while learning about choice-behavior
Dynamic pricing is example of **discrete-choice problems**:  
- Decision maker chooses attributes and/or availability of alternatives  
- Customers choose according to probabilistic choice-model  
- Decision maker aims to optimize some objective function while learning about choice-behavior  
- Many applications, expands the scope enormously
Dynamic pricing is an example of discrete-choice problems:

- Decision maker chooses attributes and/or availability of alternatives
- Customers choose according to probabilistic choice-model
- Decision maker aims to optimize some objective function while learning about choice-behavior
- Many applications, expands the scope enormously

Open PhD Position:
http://www.utwente.nl/ewi/sor/about/staff/boer/
Integrate the statistical and optimization aspects of OR problems.

- Is closer to practice
- Brings many nice theoretical challenges
Integrate the statistical and optimization aspects of OR problems.

- Is closer to practice
- Brings many nice theoretical challenges

Thanks for your attention