Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve

Ian Weaver, Vineet Kumar and Lalit Jain

Yale School of Management

Purdue University
November 2025

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets
- What is the role of theory or domain knowledge?

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets
- What is the role of theory or domain knowledge?
  - Modeling Choices: Variable (Feature) selection, Hyperparameter selection, Functional form

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets
- What is the role of theory or domain knowledge?
  - Modeling Choices: Variable (Feature) selection, Hyperparameter selection, Functional form
  - Results: Interpretation and Contextualizing

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets
- What is the role of theory or domain knowledge?
  - Modeling Choices: Variable (Feature) selection, Hyperparameter selection, Functional form
  - Results: Interpretation and Contextualizing
- Is there any role for theory (or more broadly human domain knowledge) in ML models?

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets
- What is the role of theory or domain knowledge?
  - Modeling Choices: Variable (Feature) selection, Hyperparameter selection, Functional form
  - Results: Interpretation and Contextualizing
- Is there any role for theory (or more broadly human domain knowledge) in ML models?
  - How can we incorporate (simple) economic theory?

Role of (Human) Knowledge in Models

- Machine learning models (supervised) can be cast as learning flexible functions: $\boxed{y = f(x)}$
  - Regression, SVMs, Deep Neural Nets
- What is the role of theory or domain knowledge?
  - Modeling Choices: Variable (Feature) selection, Hyperparameter selection, Functional form
  - Results: Interpretation and Contextualizing
- Is there any role for theory (or more broadly human domain knowledge) in ML models?
  - How can we incorporate (simple) economic theory?
  - What is the benefit?

The Problem - Learning Unknown Demand Curves

- In 2016, the Atlanta Falcons dramatically slashed concessions prices to improve brand equity. When asked how they projected the volume of sales to change, the CEO of ownership, Steve Cannon replied, "It could be a **10% bump**, it could be a **30% bump**, who knows." The next season sales volume for food **rose 50%**.

Setting – Pricing a New Product

- Historical/Competitor Data

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?
- Cost Analysis - calculate production price and apply a markup

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?
- Cost Analysis - calculate production price and apply a markup
  - Is this the optimal price?

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?
- Cost Analysis - calculate production price and apply a markup
  - Is this the optimal price?
- Conjoint Analysis

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?
- Cost Analysis - calculate production price and apply a markup
  - Is this the optimal price?
- Conjoint Analysis
  - Possible, but surveys are stated preferences

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?
- Cost Analysis - calculate production price and apply a markup
  - Is this the optimal price?
- Conjoint Analysis
  - Possible, but surveys are stated preferences

Setting – Pricing a New Product

- Historical/Competitor Data
  - Will that be available for a sufficiently novel product?
- Cost Analysis - calculate production price and apply a markup
  - Is this the optimal price?
- Conjoint Analysis
  - Possible, but surveys are stated preferences

Solution?

Experimentation

But... Companies are reluctant to experiment with price

Column: Why Businesses Don't Experiment

A debate ensued among the group: Are we willing to sacrifice some customers "just" to learn how the new pricing approaches work?

They hedged. They asked me what *I* thought the best approach was. I told them that I was willing to share my intuition but that intuition is a remarkably bad thing to rely on. Only an experiment gives you the evidence you need. In the end, it wasn't enough to convince them, and they called off the project.

Issues

- Potential short term losses for uncertain long term gains
- May confuse or alter consumers' expectations

But... Companies are reluctant to experiment with price

Column: Why Businesses Don't Experiment

A debate ensued among the group: Are we willing to sacrifice some customers "just" to learn how the new pricing approaches work?

They hedged. They asked me what *I* thought the best approach was. I told them that I was willing to share my intuition but that intuition is a remarkably bad thing to rely on. Only an experiment gives you the evidence you need. In the end, it wasn't enough to convince them, and they called off the project.

Issues

- Potential short term losses for uncertain long term gains
- May confuse or alter consumers' expectations

Can incorporating theory improve learning with an **unknown** demand curve?

Desired Features For Price Experimentation

Goal

Maximizing earning while learning

- Maximize long run profits while minimizing costs of experimentation

Desired Features For Price Experimentation

Goal

Maximizing earning while learning

- Maximize long run profits while minimizing costs of experimentation

Desired Features For Price Experimentation

Goal

Maximizing earning while learning

- Maximize long run profits while minimizing costs of experimentation

Desired Features

- Be able to test many prices

Desired Features For Price Experimentation

Goal

Maximizing earning while learning

- Maximize long run profits while minimizing costs of experimentation

Desired Features

- Be able to test many prices
- No parametric form for demand curve assumed

Desired Features For Price Experimentation

Goal

Maximizing earning while learning

- Maximize long run profits while minimizing costs of experimentation

Desired Features

- Be able to test many prices
- No parametric form for demand curve assumed
- Adaptively set price based on incoming experimentation data

Roadmap

Current Approach

- Multi-armed Bandits or Reinforcement Learning

Experimentation Setting

Approach

Roadmap

Current Approach

- Multi-armed Bandits or Reinforcement Learning

Experimentation Setting

- Any product with unknown demand curve (mostly new/unique)

Approach

Roadmap

Current Approach

- Multi-armed Bandits or Reinforcement Learning

Experimentation Setting

- Any product with unknown demand curve (mostly new/unique)
- The ability to change the price quickly (online setting)

Approach

Roadmap

Current Approach

- Multi-armed Bandits or Reinforcement Learning

Experimentation Setting

- Any product with unknown demand curve (mostly new/unique)
- The ability to change the price quickly (online setting)
- No inventory constraints (possible to incorporate into model)

Approach

Roadmap

Current Approach

- Multi-armed Bandits or Reinforcement Learning

Experimentation Setting

- Any product with unknown demand curve (mostly new/unique)
- The ability to change the price quickly (online setting)
- No inventory constraints (possible to incorporate into model)
- No demographic variables required (possible to incorporate into model)

Approach

Roadmap

Current Approach

- Multi-armed Bandits or Reinforcement Learning

Experimentation Setting

- Any product with unknown demand curve (mostly new/unique)
- The ability to change the price quickly (online setting)
- No inventory constraints (possible to incorporate into model)
- No demographic variables required (possible to incorporate into model)

Approach

- Combine bandits with human domain knowledge about pricing (i.e. incorporate basic economic theory)

Reinforcement Learning

Definition of RL

Reinforcement learning (RL) is a type of machine learning where an "agent" learns optimal behavior through interaction with its environment. Rather than relying on explicit programming or labeled datasets, this agent learns by trial and error, receiving feedback in the form of rewards or penalties for its actions. – Google on Reinforcement Learning

Power of RL

- Deep Blue defeated world Chess champion in 1997

Power of RL

- Deep Blue defeated world Chess champion in 1997
- AlphaGO defeated leading Go player in 2017

Power of RL

- Deep Blue defeated world Chess champion in 1997
- AlphaGO defeated leading Go player in 2017
  - How was AlphaGO trained?

Power of RL

- Deep Blue defeated world Chess champion in 1997
- AlphaGO defeated leading Go player in 2017
  - How was AlphaGO trained?
- AlphaGo Zero was introduced in 2017 in the journal Nature

Power of RL

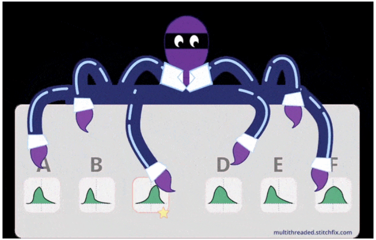
- Deep Blue defeated world Chess champion in 1997
- AlphaGO defeated leading Go player in 2017
  - How was AlphaGO trained?
- AlphaGo Zero was introduced in 2017 in the journal Nature
- AlphaGo Zero was based on RL and taught itself in 21 days

Power of RL

- Deep Blue defeated world Chess champion in 1997
- AlphaGO defeated leading Go player in 2017
    - How was AlphaGO trained?
- AlphaGo Zero was introduced in 2017 in the journal Nature
- AlphaGo Zero was based on RL and taught itself in 21 days
- Defeated AlphaGo 100–0

Introducing... Multi-Armed Bandits

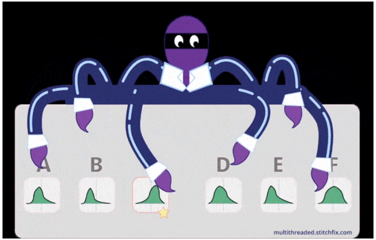
- Each arm represents a choice



Link to Animation

Introducing... Multi-Armed Bandits

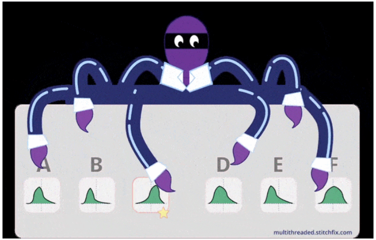
- Each arm represents a choice
  - Has unknown distribution of rewards



Link to Animation

Introducing... Multi-Armed Bandits

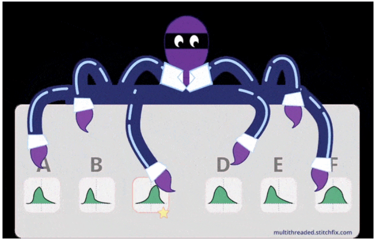
- Each arm represents a choice
  - Has unknown distribution of rewards
- Goal is to maximize rewards by choosing a sequence of arms from a set of choices (prices)



Link to Animation

Introducing... Multi-Armed Bandits

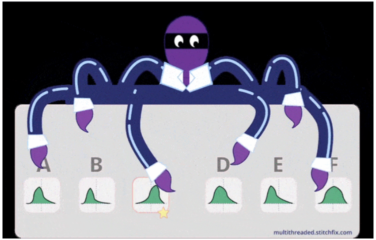
- Each arm represents a choice
  - Has unknown distribution of rewards
- Goal is to maximize rewards by choosing a sequence of arms from a set of choices (prices)
  - Learning while earning



Link to Animation

Introducing... Multi-Armed Bandits

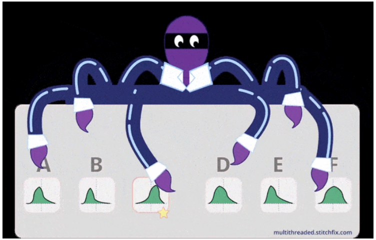
- Each arm represents a choice
  - Has unknown distribution of rewards
- Goal is to maximize rewards by choosing a sequence of arms from a set of choices (prices)
  - Learning while earning
- Key is to balance exploration (finding best arm) and exploitation (gaining from best arm)



Link to Animation

Introducing... Multi-Armed Bandits

- Each arm represents a choice
  - Has unknown distribution of rewards
- Goal is to maximize rewards by choosing a sequence of arms from a set of choices (prices)
  - Learning while earning
- Key is to balance exploration (finding best arm) and exploitation (gaining from best arm)
- Belongs to class of reinforcement learning problems



Link to Animation

Multi-Armed Bandits in Marketing

Use Case	Objective (maximize)	Examples of Arms	Arm A	Arm B	Arm C
Advertisements	Conversions	Ads	Emotional	Informative	Funny
Recommendation Systems	Purchases	Movies	E.T.	Indiana Jones	Rain Man
Pricing	**Profits**	**Prices**	**\$1**	**\$5**	**\$20**

Extra-Terrestrial? Economic Theory?

Why is Pricing Different?

Use Case	**Objective**	**Arms**	**Arm A**	**Arm B**	**Arm C**
Advertisements **Pricing**	Conversions **Profits**	Ads **Prices**	Emotional **\$1**	Informative **\$5**	Funny **\$20**

Why is Pricing Different?

Use Case	Objective	Arms	Arm A	Arm B	Arm C
Advertisements	Conversions	Ads	Emotional	Informative	Funny
Pricing	**Profits**	**Prices**	**$1**	**$5**	**$20**

In pricing, we know something about relation between arms from theory

- **Continuity:** Demands at closer prices will be closer together

Why is Pricing Different?

Use Case	Objective	Arms	Arm A	Arm B	Arm C
Advertisements	Conversions	Ads	Emotional	Informative	Funny
Pricing	**Profits**	**Prices**	**$1**	**$5**	**$20**

In pricing, we know something about relation between arms from theory

- **Continuity:** Demands at closer prices will be closer together
- **Monotonicity:** Demand is higher at lower prices

Why is Pricing Different?

Use Case	Objective	Arms	Arm A	Arm B	Arm C
Advertisements	Conversions	Ads	Emotional	Informative	Funny
Pricing	**Profits**	**Prices**	**$1**	**$5**	**$20**

In pricing, we know something about relation between arms from theory

- **Continuity:** Demands at closer prices will be closer together
- **Monotonicity:** Demand is higher at lower prices

Why is Pricing Different?

Use Case	Objective	Arms	Arm A	Arm B	Arm C
Advertisements	Conversions	Ads	Emotional	Informative	Funny
Pricing	**Profits**	**Prices**	**$1**	**$5**	**$20**

In pricing, we know something about relation between arms from theory

- **Continuity:** Demands at closer prices will be closer together
- **Monotonicity:** Demand is higher at lower prices

Informational Externality

Observing outcome for arm A helps us partially learn about outcomes for other arms

Why is Pricing Different?

Use Case	Objective	Arms	Arm A	Arm B	Arm C
Advertisements	Conversions	Ads	Emotional	Informative	Funny
Pricing	**Profits**	**Prices**	**$1**	**$5**	**$20**

In pricing, we know something about relation between arms from theory

- **Continuity:** Demands at closer prices will be closer together
- **Monotonicity:** Demand is higher at lower prices

Informational Externality

Observing outcome for arm A helps us partially learn about outcomes for other arms

Goal

Leverage informational externalities within MAB framework to improve learning

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
    - Correlation between demands at arms

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
  - Correlation between demands at arms
  - Downward sloping demand curves $\leftarrow$ **key contribution**

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
  - Correlation between demands at arms
  - Downward sloping demand curves $\leftarrow$ **key contribution**

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
  - Correlation between demands at arms
  - Downward sloping demand curves ← **key contribution**

Features

- Nonparametric

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
  - Correlation between demands at arms
  - Downward sloping demand curves $\leftarrow$ **key contribution**

Features

- Nonparametric
- Provides a measure of uncertainty (variance) for every possible price level

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
  - Correlation between demands at arms
  - Downward sloping demand curves $\leftarrow$ **key contribution**

Features

- Nonparametric
- Provides a measure of uncertainty (variance) for every possible price level
- Provides an automatic approach with no human judgement required

Contributions and Features

Contribution

Develop new principled approach that incorporates informational externalities derived from theory with nonparametric bandits

- In pricing, economic theory includes
  - Correlation between demands at arms
  - Downward sloping demand curves $\leftarrow$ **key contribution**

Features

- Nonparametric
- Provides a measure of uncertainty (variance) for every possible price level
- Provides an automatic approach with no human judgement required
- Runs in real time

Multi-armed

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\implies$ **correlation between rewards** of nearby arms

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\Longrightarrow$ **correlation between rewards** of nearby arms
  - Can allow for learning from nearby arms increasing certainty given the limited number of tests

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\implies$ **correlation between rewards** of nearby arms
  - Can allow for learning from nearby arms increasing certainty given the limited number of tests
  - We model this informational externality by adding **Gaussian processes** to standard bandit methods

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\implies$ **correlation between rewards** of nearby arms
  - Can allow for learning from nearby arms increasing certainty given the limited number of tests
  - We model this informational externality by adding **Gaussian processes** to standard bandit methods
- **Information Externality 2:** Demand curves are **monotonic (weakly decreasing)**

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\Longrightarrow$ **correlation between rewards** of nearby arms
  - Can allow for learning from nearby arms increasing certainty given the limited number of tests
  - We model this informational externality by adding **Gaussian processes** to standard bandit methods
- **Information Externality 2:** Demand curves are **monotonic (weakly decreasing)**
  - Removes erroneous non-monotonic demand curves from the GP model

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\implies$ **correlation between rewards** of nearby arms
  - Can allow for learning from nearby arms increasing certainty given the limited number of tests
  - We model this informational externality by adding **Gaussian processes** to standard bandit methods
- **Information Externality 2:** Demand curves are **monotonic (weakly decreasing)**
  - Removes erroneous non-monotonic demand curves from the GP model
  - Increases learning by reducing the possible space of demand curves

How do the informational externalities help?

Baseline: Start with standard multi-armed Bandit methods (**UCB** and **TS**)

- **Information Externality 1:** Continuity $\implies$ **correlation between rewards** of nearby arms
  - Can allow for learning from nearby arms increasing certainty given the limited number of tests
  - We model this informational externality by adding **Gaussian processes** to standard bandit methods
- **Information Externality 2:** Demand curves are **monotonic (weakly decreasing)**
  - Removes erroneous non-monotonic demand curves from the GP model
  - Increases learning by reducing the possible space of demand curves
  - We propose **new sampling method** to obtain monotonic draws from the GP

Incorporating First Informational Externality

- To model correlation between arms, can use **Gaussian process** bandits (Srinivas et al. 2009)

Incorporating First Informational Externality

- To model correlation between arms, can use **Gaussian process** bandits (Srinivas et al. 2009)

- To adapt to pricing, the only difference is that the GP is modelled at the **demand level** (rather than reward level)

Incorporating First Informational Externality

- To model correlation between arms, can use **Gaussian process** bandits (Srinivas et al. 2009)
- To adapt to pricing, the only difference is that the GP is modelled at the **demand level** (rather than reward level)
  - y variable: purchase probability

Incorporating First Informational Externality

- To model correlation between arms, can use **Gaussian process** bandits (Srinivas et al. 2009)
- To adapt to pricing, the only difference is that the GP is modelled at the **demand level** (rather than reward level)
  - y variable: purchase probability
  - x variable: price

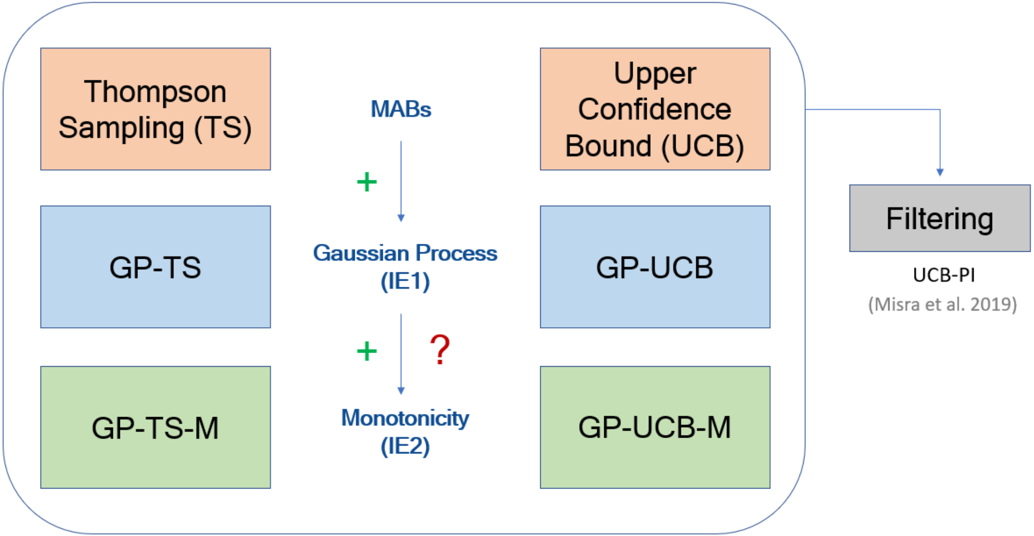
Incorporating First Informational Externality

- To model correlation between arms, can use **Gaussian process** bandits (Srinivas et al. 2009)
- To adapt to pricing, the only difference is that the GP is modelled at the **demand level** (rather than reward level)
  - y variable: purchase probability
  - x variable: price
- Information from posterior GP scaled by price accordingly (GP-UCB / GP-TS)

Building Blocks

- Decision Rule for Picking Arm to Experiment: UCB, TS
- Gaussian Process can be used to flexibly model any nonparametric demand curve
    - Monotonicity restricts the set of functions

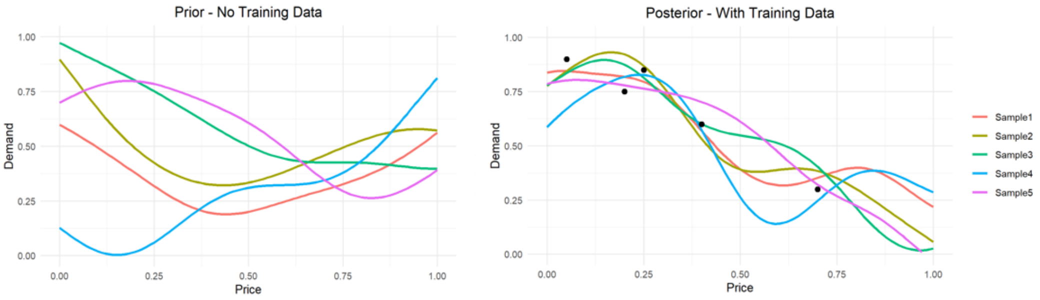
Overview of MABs and Informational Externalities



Overview of GPs

Intuitively, a Gaussian process is a probability distribution over possible functions

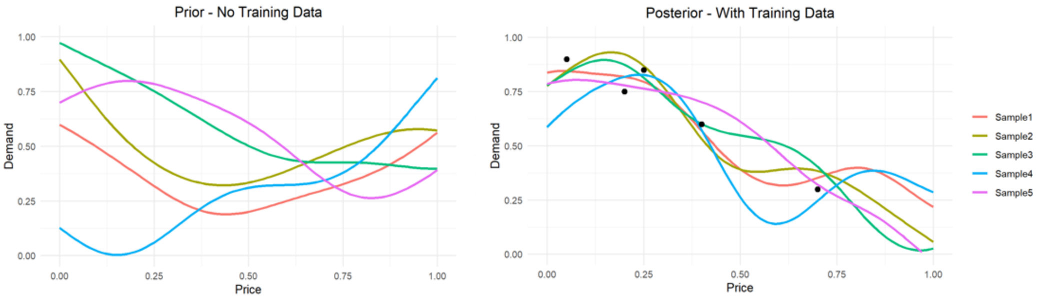
- GPs have a simple closed form specified completely by mean and kernel
  $f \sim GP(\mu(p), k(p, p'))$



Overview of GPs

Intuitively, a Gaussian process is a probability distribution over possible functions

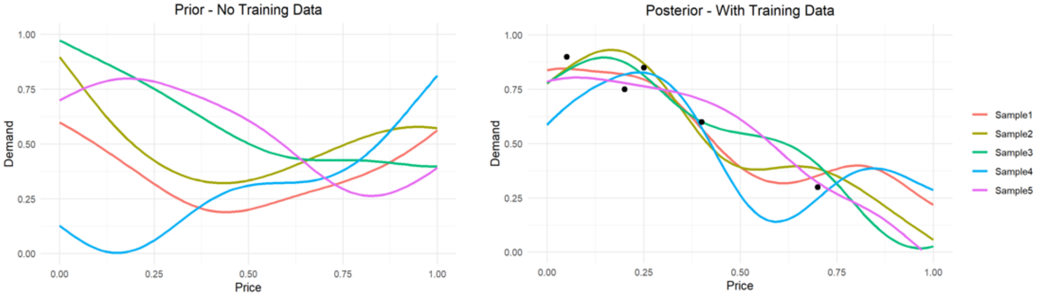
- GPs have a simple closed form specified completely by mean and kernel $f \sim GP(\mu(p), k(p, p'))$
- Nonparametric method (accommodates almost any possible demand curve)



Overview of GPs

Intuitively, a Gaussian process is a probability distribution over possible functions

- GPs have a simple closed form specified completely by mean and kernel
  $f \sim GP(\mu(p), k(p, p'))$
- Nonparametric method (accommodates almost any possible demand curve)
- Can **sample an entire demand curve** at once from GP



Incorporating Second Informational Externality

GP-UCB

Uses deterministic scoring rule to pick an arm by balancing exploitation (mean of GP) with exploration (variance of GP) for each arm

Incorporating Second Informational Externality

GP-UCB

Uses deterministic scoring rule to pick an arm by balancing exploitation (mean of GP) with exploration (variance of GP) for each arm To include monotonicity:

- Need to model GP such that the means are monotonically decreasing

GP-TS

Uses random draw from posterior, picking the arm with the highest reward
To include monotonicity

Incorporating Second Informational Externality

GP-UCB

Uses deterministic scoring rule to pick an arm by balancing exploitation (mean of GP) with exploration (variance of GP) for each arm To include monotonicity:

- Need to model GP such that the means are monotonically decreasing
- Available methods (Riihimaki & Vehtari (2010))

GP-TS

Uses random draw from posterior, picking the arm with the highest reward
To include monotonicity

Incorporating Second Informational Externality

GP-UCB

Uses deterministic scoring rule to pick an arm by balancing exploitation (mean of GP) with exploration (variance of GP) for each arm To include monotonicity:

- Need to model GP such that the means are monotonically decreasing
- Available methods (Riihimaki & Vehtari (2010))

GP-TS

Uses random draw from posterior, picking the arm with the highest reward
To include monotonicity

- Need way to obtain a random monotonically decreasing draw from the posterior GP

Incorporating Second Informational Externality

GP-UCB

Uses deterministic scoring rule to pick an arm by balancing exploitation (mean of GP) with exploration (variance of GP) for each arm To include monotonicity:

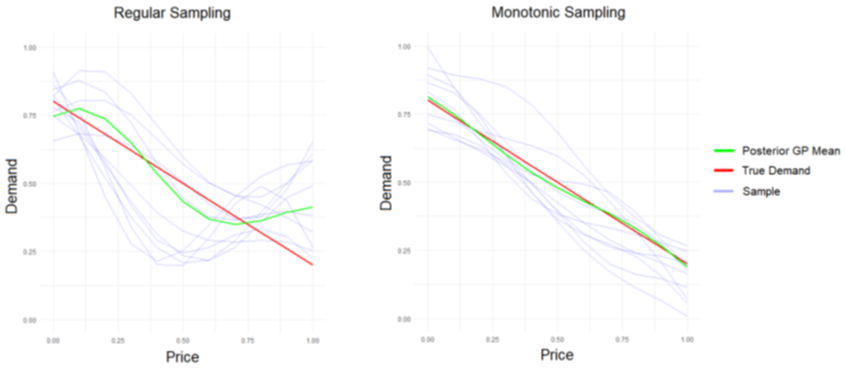
- Need to model GP such that the means are monotonically decreasing
- Available methods (Riihimaki & Vehtari (2010))

GP-TS

Uses random draw from posterior, picking the arm with the highest reward
To include monotonicity

- Need way to obtain a random monotonically decreasing draw from the posterior GP
- Not entirely clear how to do so

Monotonic Sampling



How do we obtain a random monotonic draw from the GP in a principled way?

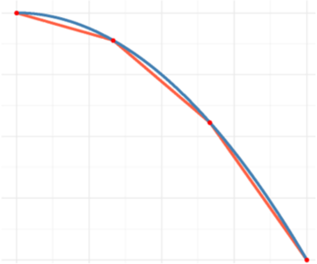
Proposed Sampling Approach (I)

Proposed Approach

Any function can be estimated by linearly interpolating between knot points

- Converges to true curve as number of knots approaches infinity

$$D(\cdot) \approx \sum_{j=0}^{N} D(\mu_j) h_j(\cdot)$$



Similarly, any function can be estimated by its intercept and first derivatives at knots

$$D(p) \approx \sum_{j=0}^{N} D'(\mu_j) \int_0^p h_j(t)dt$$

Proposed Sampling Approach (II)

Proposed Approach

[1]We use the *TruncatedNormal* package in R

Proposed Sampling Approach (II)

Proposed Approach

- Demand function can be estimated in terms of its intercept and derivatives at various knots (prices)
- Possible because the sum and derivative of a GP is also a GP

Why would we want to do this?

- Transforms problem from one of obtaining a monotonic draw to one of obtaining negative draws

[1] We use the *TruncatedNormal* package in R

Proposed Sampling Approach (II)

Proposed Approach

- Demand function can be estimated in terms of its intercept and derivatives at various knots (prices)
- Possible because the sum and derivative of a GP is also a GP

Why would we want to do this?

- Transforms problem from one of obtaining a monotonic draw to one of obtaining negative draws
- Monotonically decreasing = all negative derivatives

[1] We use the *TruncatedNormal* package in R

Proposed Sampling Approach (II)

Proposed Approach

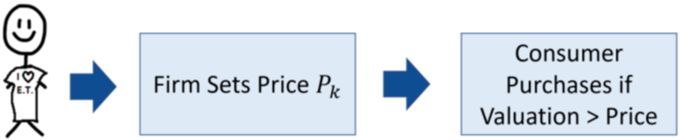
- Demand function can be estimated in terms of its intercept and derivatives at various knots (prices)
- Possible because the sum and derivative of a GP is also a GP

Why would we want to do this?

- Transforms problem from one of obtaining a monotonic draw to one of obtaining negative draws
- Monotonically decreasing = all negative derivatives
- Easier sampling problem (that of sampling from a Truncated Normal[1])

[1] We use the *TruncatedNormal* package in R

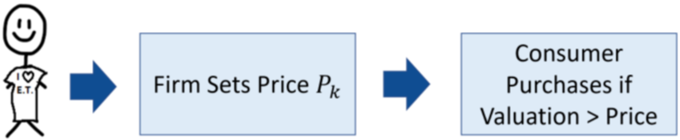
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen

[2] Follows setup from Misra et al. (2019)

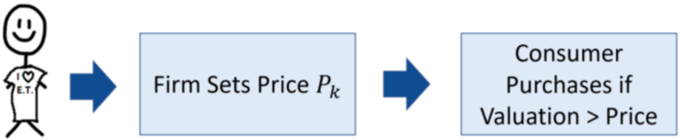
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1

[2] Follows setup from Misra et al. (2019)

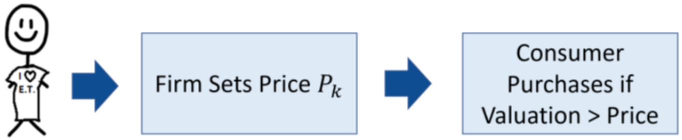
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1
- 3 different price sets are tried

[2] Follows setup from Misra et al. (2019)

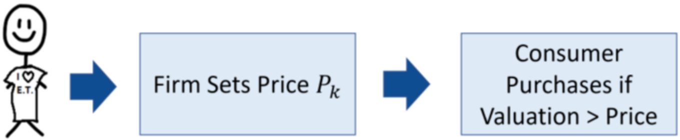
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1
- 3 different price sets are tried
  - 5 arms: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$

[2] Follows setup from Misra et al. (2019)

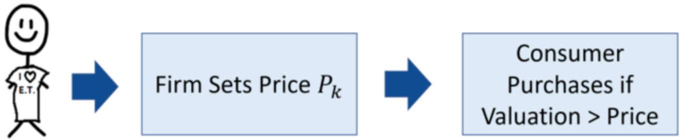
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1
- 3 different price sets are tried
  - 5 arms: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$
  - 10 arms: $\{0.1, 0.2, ..., 1\}$

[2] Follows setup from Misra et al. (2019)

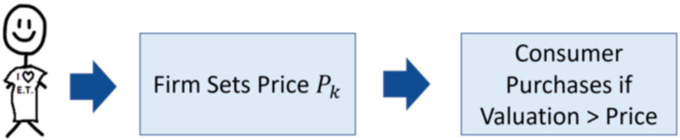
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1
- 3 different price sets are tried
  - 5 arms: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$
  - 10 arms: $\{0.1, 0.2, ..., 1\}$
  - 100 arms: $\{0.01, 0.02, ..., 1\}$

[2]Follows setup from Misra et al. (2019)

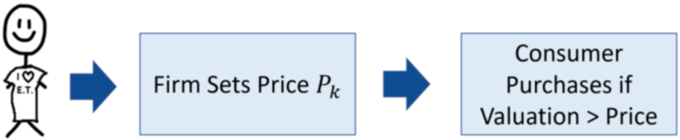
Algorithm Evaluation – Setup



- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1
- 3 different price sets are tried
    - 5 arms: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$
    - 10 arms: $\{0.1, 0.2, ..., 1\}$
    - 100 arms: $\{0.01, 0.02, ..., 1\}$
- Algorithm is updated with purchase decision every 10 customers[2]

[2] Follows setup from Misra et al. (2019)

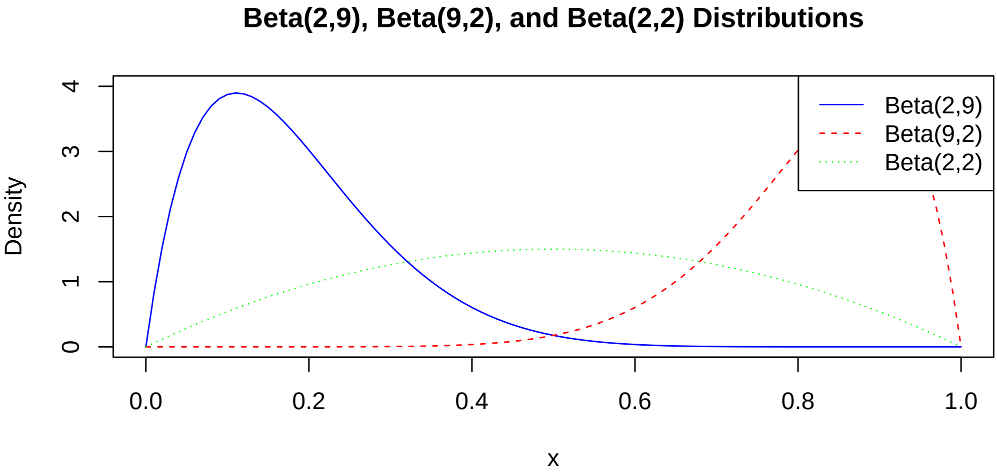
Algorithm Evaluation – Setup



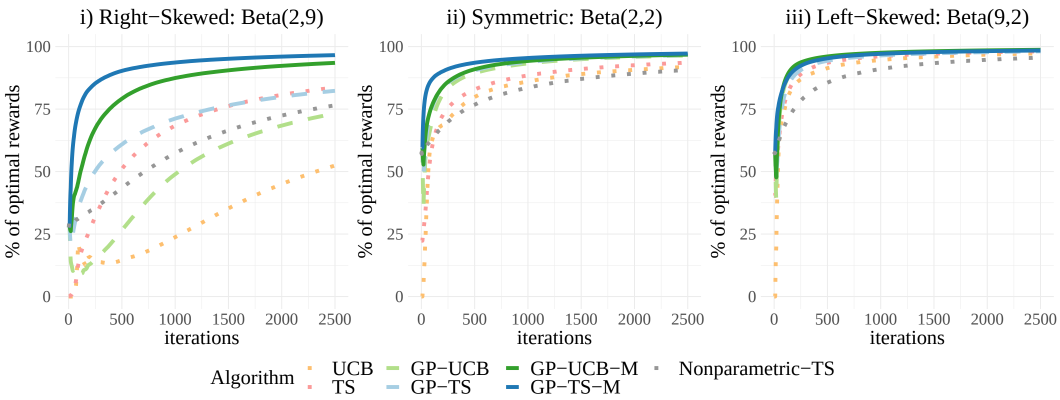
- Underlying WTP distribution is chosen
- Prices are normalized between 0 and 1
- 3 different price sets are tried
  - 5 arms: $\{0.1, 0.3, 0.5, 0.7, 0.9\}$
  - 10 arms: $\{0.1, 0.2, ..., 1\}$
  - 100 arms: $\{0.01, 0.02, ..., 1\}$
- Algorithm is updated with purchase decision every 10 customers[2]
- Results are averaged across 1000 separate runs

[2] Follows setup from Misra et al. (2019)

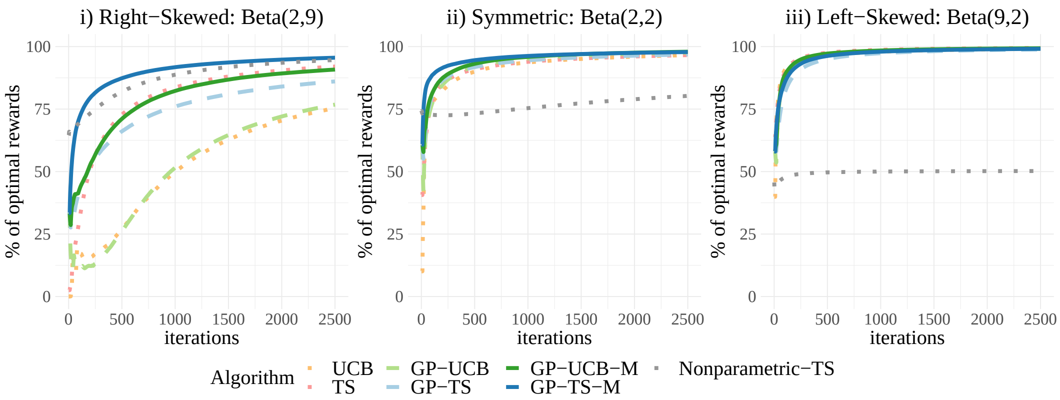
Distributions



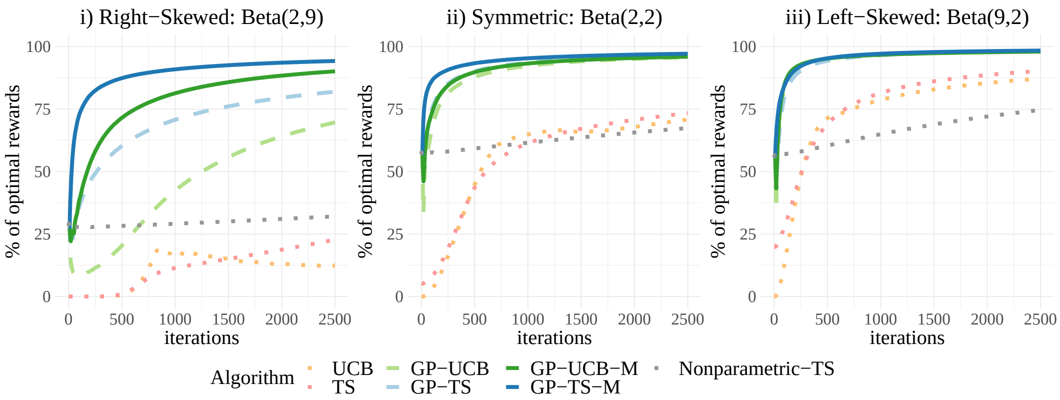
Rewards - 10 Arms



Rewards - 5 Arms



Rewards - 100 Arms



Performance Uplift from Informational Externalities

2500 consumers

After 2500 Consumers

		TS			UCB		
		5 Arms	10 Arms	100 Arms	5 Arms	10 Arms	100 Arms
Uplift from 1st externality (GP compared to base algos)	**B(2,9)**	-6.4% (-6.5, -6.2)	-1.6% (-1.9, -1.3)	262% (260, 264)	2.2% (1.9, 2.4)	40.4% (39.7, 41.1)	464% (462, 467)
	B(2,2)	0.3% (0.1, 0.4)	3.4% (3.2, 3.5)	31.4% (31.2, 31.6)	1.0% (0.9, 1.1)	5.0% (4.8, 5.1)	35.2% (35.0, 35.4)
	B(9,2)	-0.5% (-0.6, -0.5)	0.0% (0.0, 0.1)	8.8% (8.7, 8.9)	0.1% (0.0, 0.1)	1.3% (1.3, 1.4)	12.5% (12.4, 12.6)
	Mean	**-2.1% (-2.2, -2.0)**	**0.6% (0.5, 0.8)**	**48.4% (48.2, 48.6)**	**1.0% (0.9, 1.1)**	**11.1% (11.0, 11.3)**	**54.7% (54.5, 54.9)**
Uplift from 2nd externality (GP-M compared to GP algos)	**B(2,9)**	10.9% (10.6, 11.1)	17.4% (17.1, 17.8)	14.9% (14.6, 15.2)	18.0% (17.6, 18.5)	27.6% (27.1, 28.1)	29.4% (28.9, 30.0)
	B(2,2)	1.0% (0.9, 1.1)	0.7% (0.5, 0.8)	0.9% (0.7, 1.0)	0.4% (0.3, 0.5)	0.3% (0.2, 0.5)	0.3% (0.1, 0.4)
	B(9,2)	0.2% (0.2, 0.3)	0.3% (0.2, 0.3)	0.3% (0.2, 0.3)	0.1% (0.1, 0.2)	0.2% (0.1, 0.2)	-0.1% (-0.1, 0.0)
	Mean	**3.7% (3.6, 3.8)**	**5.5% (5.4, 5.6)**	**4.8% (4.7, 4.9)**	**5.3% (5.1, 5.4)**	**7.7% (7.5, 7.8)**	**7.8% (7.7, 7.9)**
Uplift from both externalities (GP-M compared to base algos)	**B(2,9)**	3.8% (3.6, 4.0)	15.4% (15.2, 15.6)	315% (313, 318)	20.5% (20.0, 21.0)	78.9% (78.1, 79.7)	629% (627, 632)
	B(2,2)	1.3% (1.1, 1.4)	4.0% (3.9, 4.2)	32.5% (32.3, 32.7)	1.4% (1.3, 1.6)	5.3% (5.1, 5.5)	35.6% (35.4, 35.8)
	B(9,2)	-0.3% (-0.4, -0.3)	0.3% (0.2, 0.4)	9.0% (8.9, 9.1)	0.2% (0.2, 0.3)	1.5% (1.4, 1.5)	12.4% (12.3, 12.5)
	Mean	**1.5% (1.4, 1.6)**	**6.2% (6.1, 6.3)**	**55.5% (55.4, 55.7)**	**6.3% (6.2, 6.4)**	**19.6% (19.5, 19.8)**	**66.8% (66.6, 66.9)**

1st Externality

- Uplift increases with number of arms
- Magnitude of uplifts largest for right-skewed

2nd Externality

- Consistent uplifts regardless of number of arms
- Large uplift for right-skewed, small for symmetric and left-skewed

Performance Uplift from Informational Externalities

500 consumers

		After 500 Consumers					
		TS			UCB		
		5 Arms	10 Arms	100 Arms	5 Arms	10 Arms	100 Arms
Uplift from 1st externality (GP compared to base algos)	B(2,9)	-8.0% (-8.6, -7.3)	20.6% (19.2, 22.1)	5940% (5840, 6040)	2.0% (0.0, 4.1)	92.1% (86.6, 97.6)	2720% (2650, 2790)
	B(2,2)	1.0% (0.6, 1.3)	9.4% (9.0, 9.8)	108% (107, 109)	2.4% (2.0, 2.7)	12.6% (12.1, 13.1)	96.6% (96.0, 97.2)
	B(9,2)	-2.4% (-2.5, -2.2)	0.8% (0.6, 1.0)	36.8% (36.6, 37.1)	0.1% (0.0, 0.3)	4.6% (4.5, 4.8)	32.7% (32.6, 32.8)
	Mean	**-2.8% (-3.1, -2.6)**	**8.2% (7.8, 8.5)**	**116% (115, 117)**	**0.9% (0.7, 1.2)**	**14.3% (13.9, 14.7)**	**73.9% (73.4, 74.4)**
Uplift from 2nd externality (GP-M compared to GP algos)	B(2,9)	31.5% (30.5, 32.5)	50.1% (48.4, 51.8)	45.5% (44.1, 47.0)	176% (170, 182)	237% (222, 252)	296% (281, 310)
	B(2,2)	2.9% (2.6, 3.2)	3.5% (3.2, 3.9)	3.4% (3.1, 3.8)	1.3% (1.0, 1.6)	1.4% (1.0, 1.8)	2.1% (1.6, 2.5)
	B(9,2)	1.0% (0.9, 1.2)	1.1% (1.0, 1.2)	1.1% (0.9, 1.2)	0.6% (0.5, 0.7)	0.5% (0.4, 0.6)	0.5% (0.4, 0.5)
	Mean	**9.6% (9.3, 9.8)**	**13.7% (13.3, 14.0)**	**12.6% (12.3, 12.9)**	**21.5% (21.1, 21.9)**	**25.6% (25.1, 26.0)**	**26.3% (25.9, 26.7)**
Uplift from both externalities (GP-M compared to base algos)	B(2,9)	20.4% (19.6, 21.3)	78.5% (76.9, 80.0)	8610% (8470, 8740)	174% (168, 179)	468% (457, 479)	975% (970, 980)
	B(2,2)	3.8% (3.5,4.1)	13.2% (12.7, 13.6)	115% (114,116)	3.6% (3.2, 3.9)	14.1% (13.6, 14.5)	100% (100, 101)
	B(9,2)	-1.4% (-1.5, -1.2)	1.9% (1.7, 2.1)	38.3% (38.0, 38.6)	0.8% (0.7, 0.9)	5.2% (5.0, 5.3)	33.3% (33.2, 33.4)
	Mean	**6.4% (6.1, 6.6)**	**22.9% (22.5, 23.2)**	**143% (143, 144)**	**22.6% (22.1, 23.0)**	**43.3% (42.9, 43.7)**	**119% (119, 120)**

1st Externality

- Uplift increases with number of arms
- Magnitude of uplifts largest for right-skewed

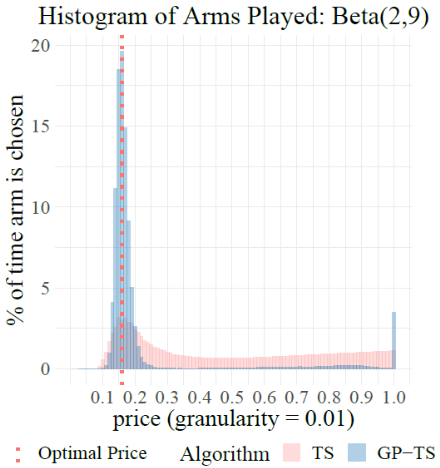
2nd Externality

- Consistent uplifts regardless of number of arms
- Large uplift for right-skewed, small for symmetric and left-skewed

Explanation - 1st Externality

The first externality becomes more important as the number of arms increases

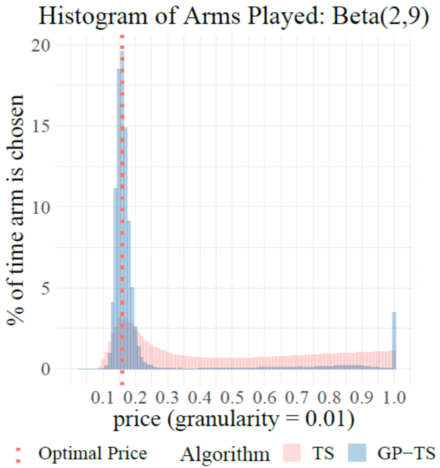
- Without considering correlation, an algorithm has to test each arm individually



Explanation - 1st Externality

The first externality becomes more important as the number of arms increases

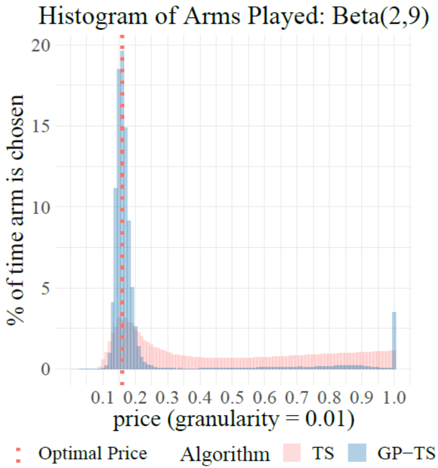
- Without considering correlation, an algorithm has to test each arm individually
- As the number of arms increases, learning is spread too thin



Explanation - 1st Externality

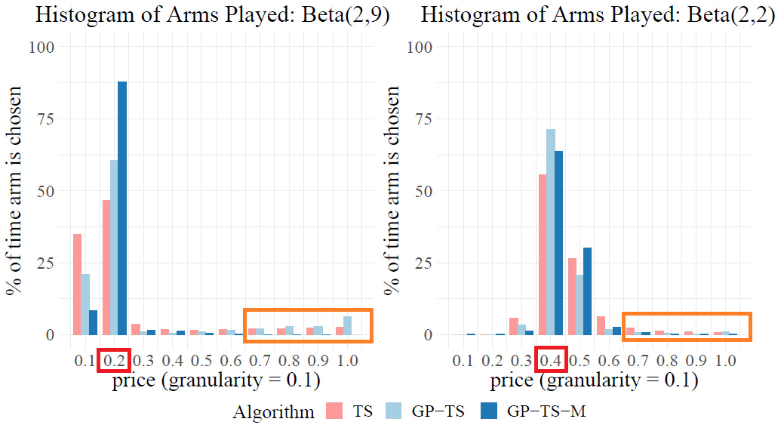
The first externality becomes more important as the number of arms increases

- Without considering correlation, an algorithm has to test each arm individually
- As the number of arms increases, learning is spread too thin
- Learning across arms reduces this problem allowing algorithm to narrow in on best arms more quickly



Explanation - 2nd Externality

- Generally, bandit algorithms tend to over-explore higher prices

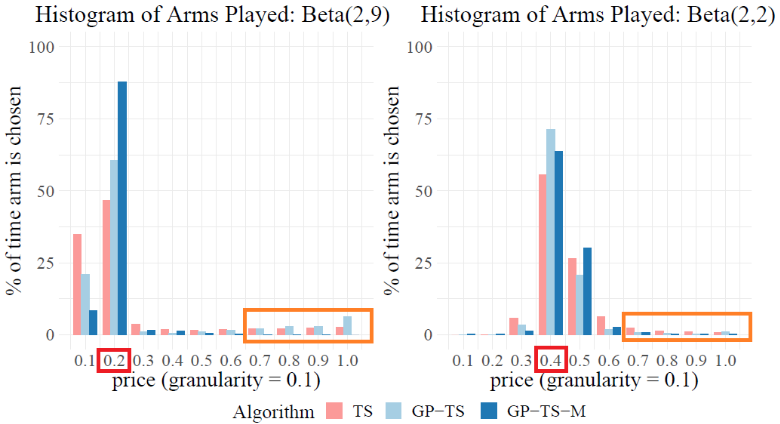


- For Beta(2,9), including monotonicity helps remove the choice of many low performing high prices (orange box)

- For Beta(2,2), the effect is largely mitigated because other algorithms are already learning quickly to move on from these low performing high prices (orange box)

Explanation - 2nd Externality

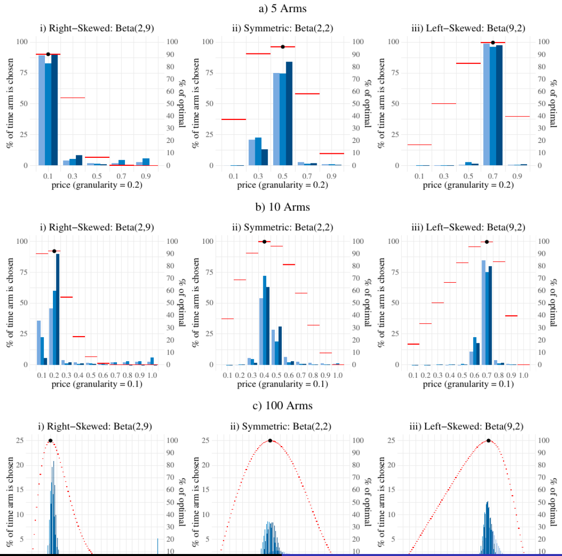
- Generally, bandit algorithms tend to over-explore higher prices
- When the optimal price is low, including monotonicity allows many 'exploitations' from erroneous upward sloping curves to be eliminated



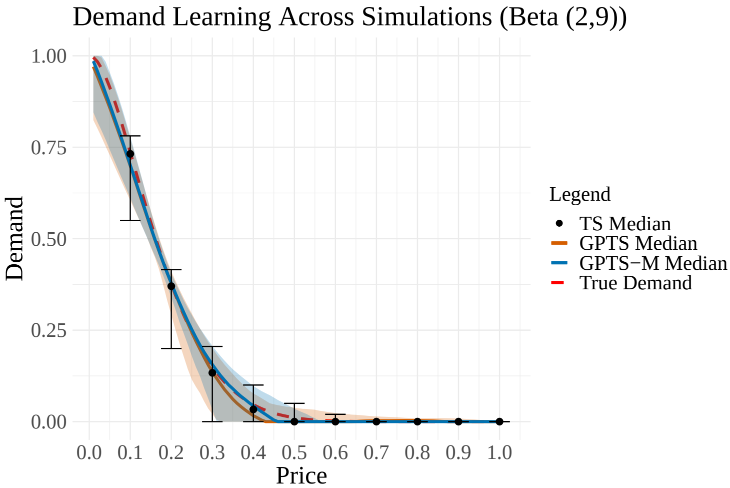
- For Beta(2,9), including monotonicity helps remove the choice of many low performing high prices (orange box)

- For Beta(2,2), the effect is largely mitigated because other algorithms are already learning quickly to move on from these low performing high prices (orange box)

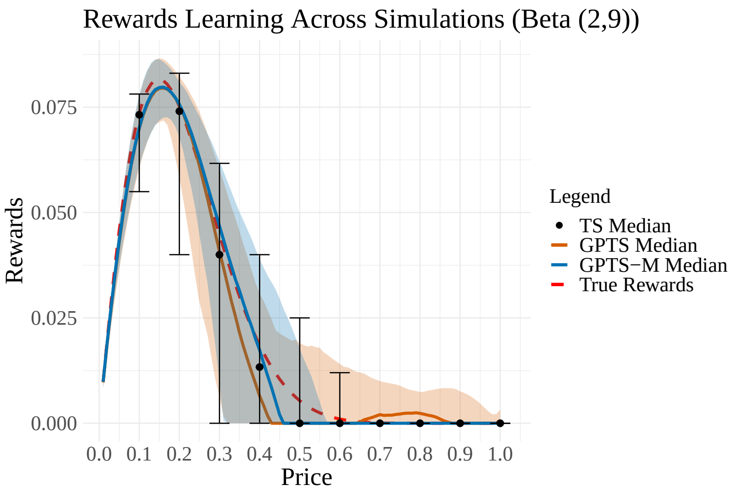
Mechanism and Explanation



Comparison of Learning Across Simulations



Comparison of Learning Across Simulations



Conclusion

- Role of Theory (Human Domain Knowledge) in Machine Learning models

Conclusion

- Role of Theory (Human Domain Knowledge) in Machine Learning models
- Incorporate "Demand Curves are continuous and downward sloping" into Reinforcement Learning (MAB) model

Conclusion

- Role of Theory (Human Domain Knowledge) in Machine Learning models
- Incorporate "Demand Curves are continuous and downward sloping" into Reinforcement Learning (MAB) model
  - Want to retain flexibility of ML models (Nonparametric, incredible flexible forms)

Conclusion

- Role of Theory (Human Domain Knowledge) in Machine Learning models
- Incorporate "Demand Curves are continuous and downward sloping" into Reinforcement Learning (MAB) model
  - Want to retain flexibility of ML models (Nonparametric, incredible flexible forms)
- Theory $\Longrightarrow$ **Informational externalities** across arms

Conclusion

- Role of Theory (Human Domain Knowledge) in Machine Learning models
- Incorporate "Demand Curves are continuous and downward sloping" into Reinforcement Learning (MAB) model
  - Want to retain flexibility of ML models (Nonparametric, incredible flexible forms)
- Theory $\implies$ **Informational externalities** across arms
- Incorporating Theory into ML models requires careful modeling but has significant benefits in improving ML

Conclusion

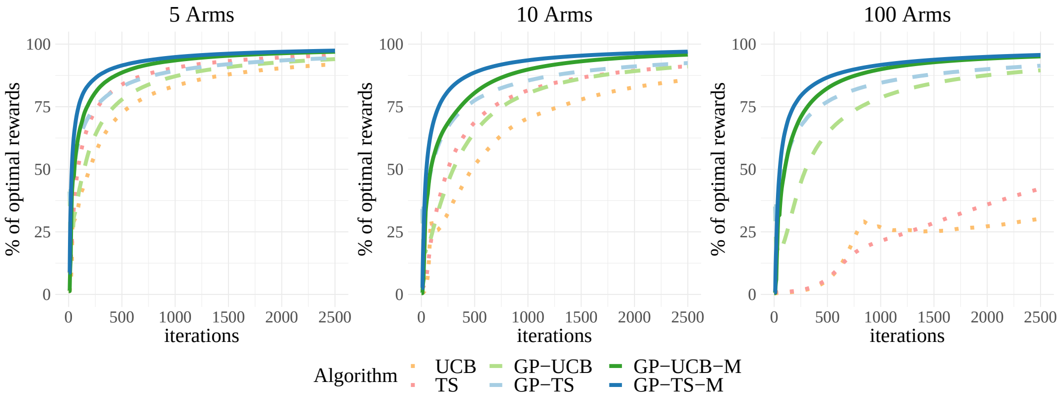
- Role of Theory (Human Domain Knowledge) in Machine Learning models
- Incorporate "Demand Curves are continuous and downward sloping" into Reinforcement Learning (MAB) model
    - Want to retain flexibility of ML models (Nonparametric, incredible flexible forms)
- Theory $\implies$ **Informational externalities** across arms
- Incorporating Theory into ML models requires careful modeling but has significant benefits in improving ML
- Learns much more *efficiently* $\implies$ Experimentation ($\Downarrow$) $\implies$ Practical value

Thank You!

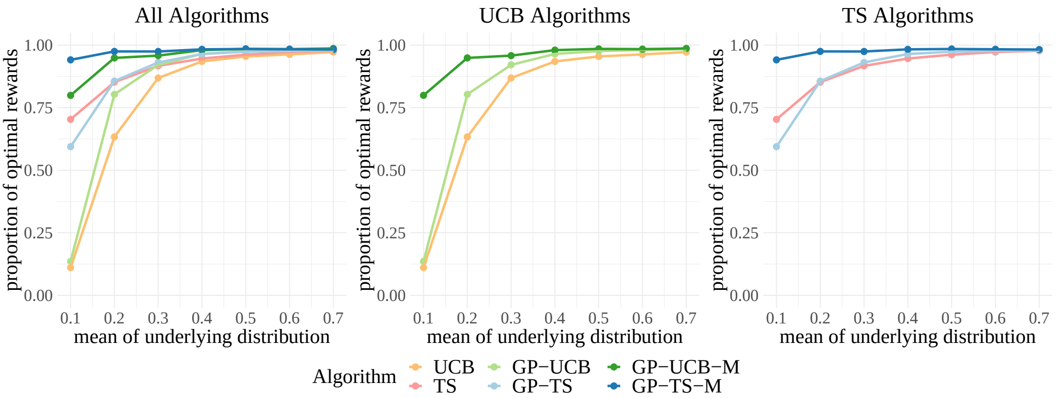
vineet.kumar@yale.edu

ADDITIONAL

Field Data - Replication



Fixed Variance - Altering the Mean of Underlying Distribution



The difference in performance shrinks as the mean gets higher