Dynamic Pricing of Ancillaries using Reinforcement Learning

AGIFORS RM Study Group Meeting, May 4-6, 2021

Ravi Kumar: PROS
(Joint work with Guangrui Xie, Ang Li and Yan Xu: PROS)

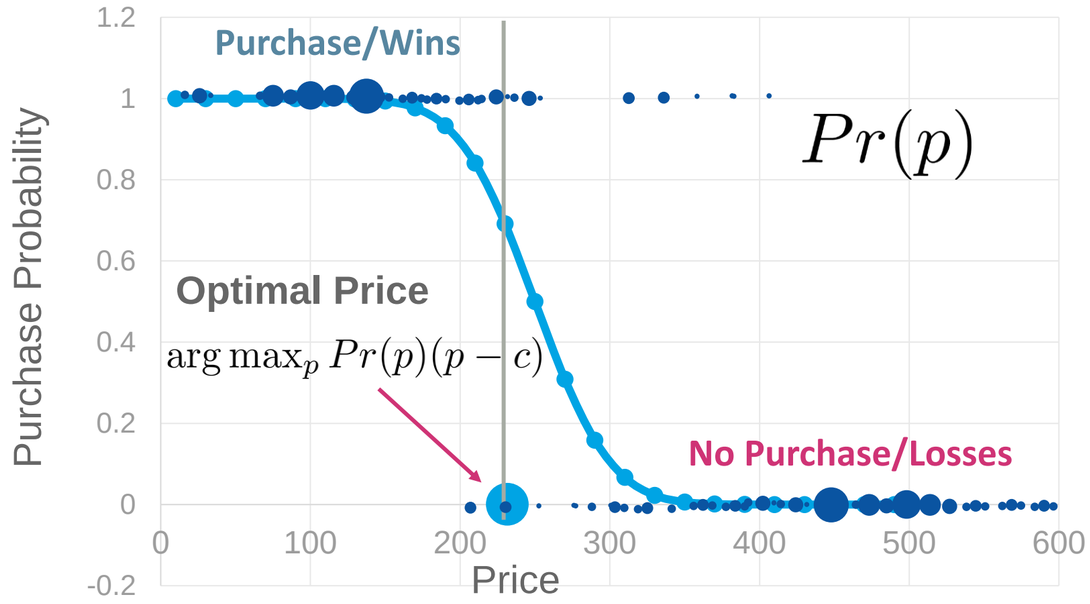
Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

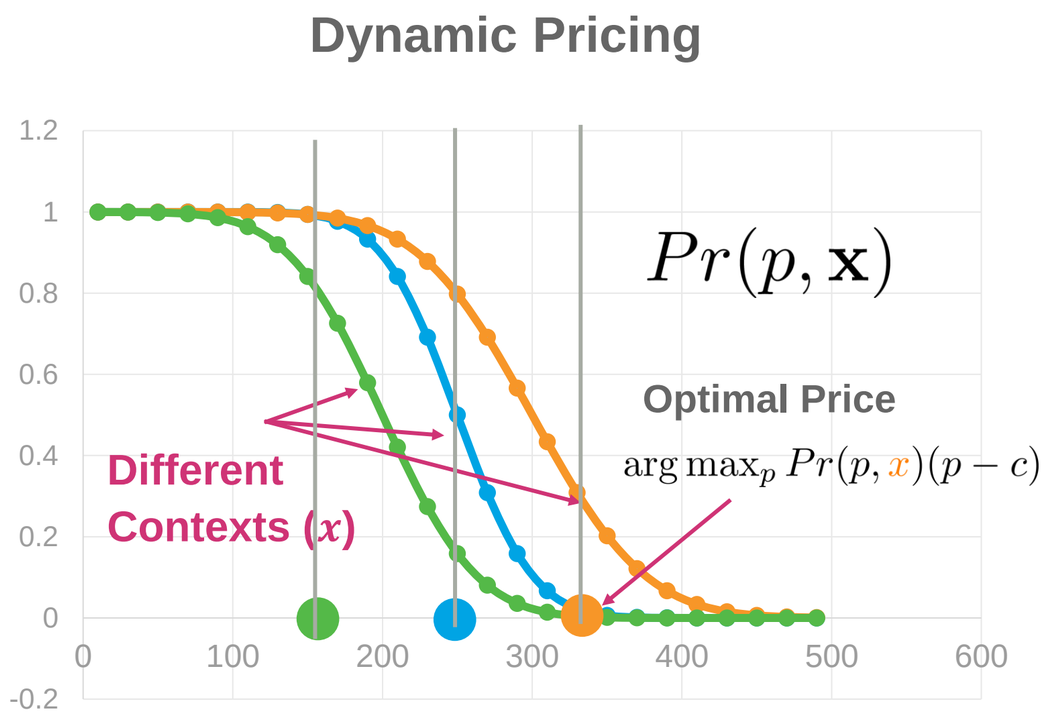
Bio

Ravi Kumar is a Lead Scientist at PROS Inc, where he has worked for over 5 years. His work involves research and development of control, estimation and optimization algorithms for Revenue Management and Dynamic Pricing systems. His research also focuses on reinforcement learning and contextual bandits for personalized recommendations. Ravi received his Ph.D. in Operations Research from Cornell University in 2015

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Dynamic Pricing Models: General Framework



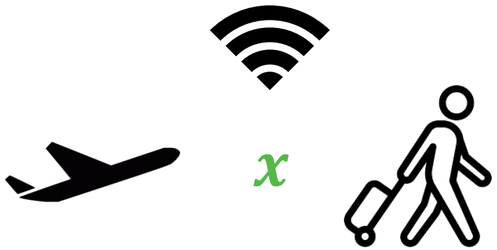


Context $\boldsymbol{x}$: product features, competitor price, events, customer features….

Traditionally dynamic pricing is driven by purchase probability functions which need to be estimated based on historical sales transaction data
Reliable estimation requires that item has been historically offered at different price points

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Dynamic Pricing with Learning



How should the seller price to quickly discover optimal prices while not loosing too much revenue?

Price based on online context x

- Consider situations where we do not have sufficiently rich history to estimate purchase probability
  - Introduced a new item
  - Historically a static pricing policy was used
  - Environment has suddenly changed e.g., pandemic and therefore consumer preferences have shifted
- The seller still wants to compute revenue optimal prices for different contexts
  - Willing to conduct “limited” price experimentation

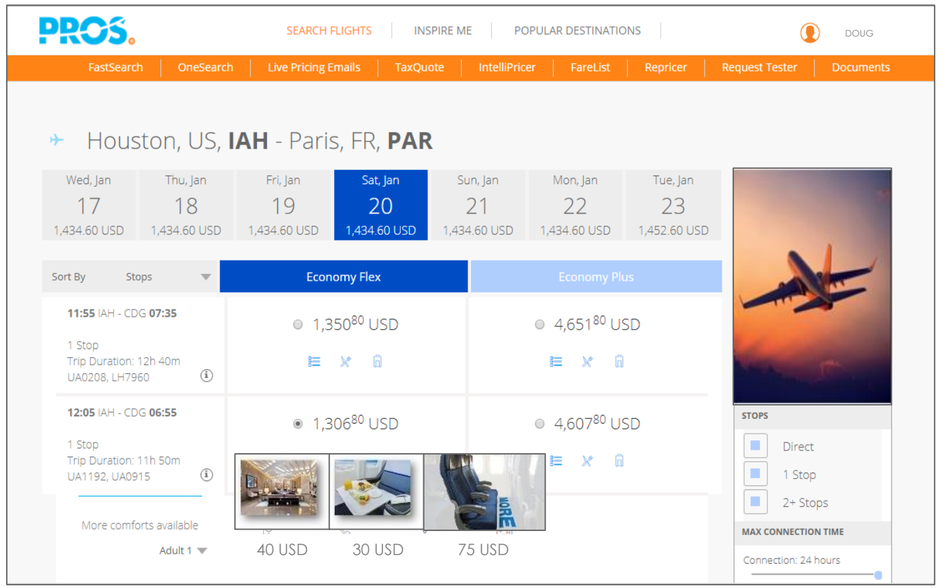
Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Airline Ancillary Pricing

Customer selects an itinerary

IAH-PAR, 20 January 2022

At subsequent stage of booking process airline offers ancillaries



Good quality loss/no purchase information available: makes estimation of purchase probability easier

Airlines can offer a variety of ancillary services

Extra Bags, Seat Selection, Lounge, Fast Access, Meals, Beverages, Specialty Equipment etc.

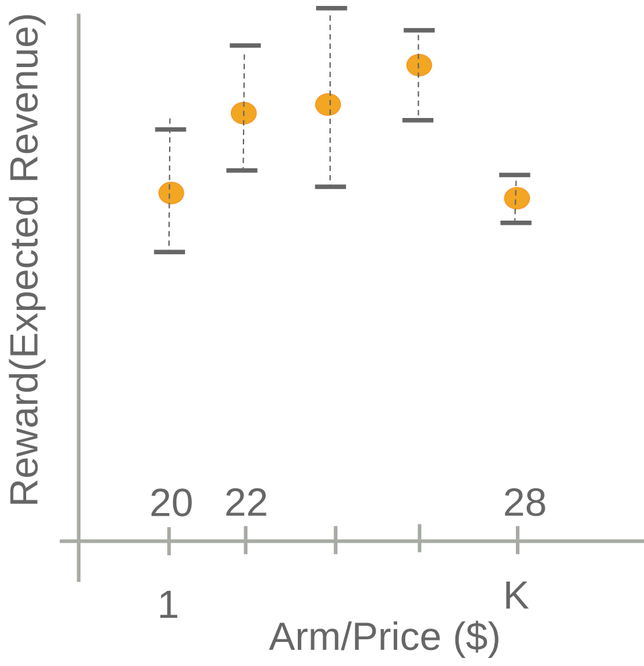
Static pricing policy implemented historically

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Dynamic Pricing with Learning: Literature Review

- Infinite Inventory Pricing Setting (continuous action set)
  - **Single context**: Araman and Caldentey (2009), Besbes and Zeevi (2009), Farias and Van Roy (2010), Harrison et al. (2012), Broder and Rusmevichientong (2012), den Boer and Zwart (2014), Keskin and Zeevi (2014, 2016), Cheung et al. (2017), Ferreira et al. (2018)
  - **Contextual Setting**: Chen et al. (2015), Cohen et al. (2016), Qiang and Bayati (2016), Javanmard and Nazerzadeh (2016), Qu et al. (2019) – Bayesian framework, Bayes-greedy i.e., myopic optimal policy

- Multi-armed Bandits (discrete action set)
  - Arms are discrete prices and have a correlated reward structure
  - Upper Confidence Bound (UCB) Heuristic
    - Achieves optimal regret ($O(\ln T)$) for independent arm single context case: Lai and Robbins (1985)
  - Thompson Sampling Heuristic
    - Achieves optimal regret ($O(\ln T)$) for independent arm single context case: Agrawal and Goyal (2012)
  - Contextual setting: Auer (2003), Goldenshluger and Zeevi (2013), Bastani and Bayati (2015)

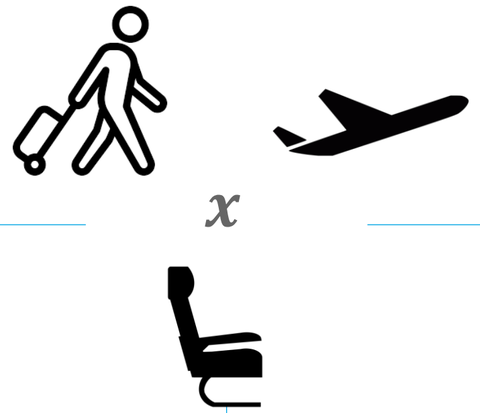


We consider a Bayesian framework (Logistic model) like Qu et al. but implement MAB heuristics for the specific case of Airline ancillary pricing

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Context Dependent Ancillary Pricing – Data Available

Traveler
FF Tier Level
Miles balance



Trip
Origin Airport
Destination Airport
Path
Departure Date
Departure Time
Booking Date
Compartment
Fare Family
Trip Way
POS
Channel

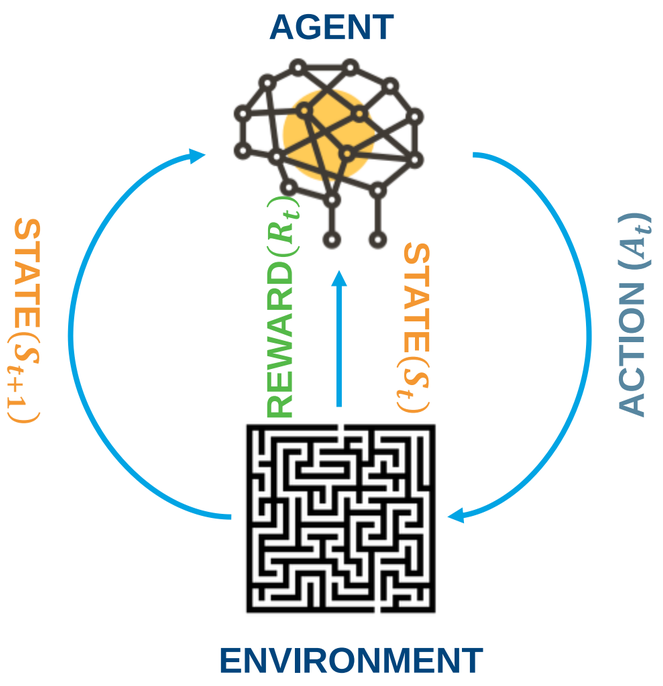
Derived Features
Origin City
Origin Country
Origin Region
Destination City
Destination Country
Destination Region
Trip Distance
Days Prior
Stay Duration
Number of Accompanying Passengers
Indicator of Intercontinental Flight

Product Features
Ancillary type: travel service, inflight service etc
Historical conversion rate
Item popularity score

Context vector for defined by the unique combination of Trip, Product features and Traveler (x_c)

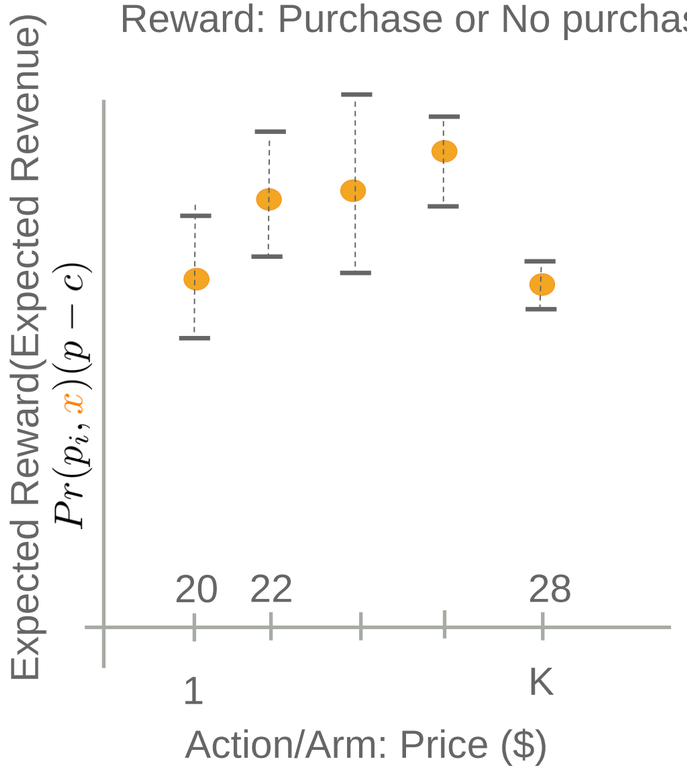
Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Reinforcement Learning and Multi-armed Bandits for Pricing



REINFORCEMENT LEARNING

State: Shopping Context (x)
Reward: Purchase or No purchase



CONTEXTUAL MULTI-ARMED BANDITS (MAB) FOR PRICING

If a customer didn't purchase at some price, they won't purchase at any higher price and vice-versa

REWARDS ACROSS ARMS IN PRICING MABs ARE CORRELATED

EXPLORE/EXPLOIT TRADEOFF

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Contextual Bandits and Pricing

- Standard linear contextual bandit approaches that are widely used in the recommender system community doesn't work well for pricing

- More advanced versions like Hybrid Linear Models [Li et al 2010] with price related engineered features provide some improvement but still not the best

$$\mathbf{E}[r_{t,a}|\mathbf{x}_{t,a}] \quad = \quad \mathbf{z}_{t,a}^{\top}\boldsymbol{\beta}^{*} + \mathbf{x}_{t,a}^{\top}\boldsymbol{\theta}_{a}^{*}$$

$\mathbf{z}_{t,a}^{\top}$: price related features p, p^2 to be modeled globally

$\boldsymbol{\beta}^{*}$: parameters of global linear model

$\mathbf{x}_{t,a}^{\top}$: price-independent features (context) within each arm

$\boldsymbol{\theta}_{a}^{*}$: parameters of local linear model within each arm

Unable to capture the special structure, reward correlation, of the pricing problem

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Logistic Model

- Observed data: Purchase or No purchase, data distribution (Bernoulli)

$$y|\boldsymbol{x},\boldsymbol{\beta} = \begin{cases} 1, w.p. & \sigma(\boldsymbol{x},\boldsymbol{\beta}) \\ 0, w.p. & 1-\sigma(\boldsymbol{x},\boldsymbol{\beta}) \end{cases}$$

- Purchase probability governed by logistic function

$$\sigma(\boldsymbol{x},\boldsymbol{\beta}) = \frac{1}{1+e^{\boldsymbol{\beta}^T\boldsymbol{x}}}$$

- $\boldsymbol{x}$ is built by augmenting the context vector $\boldsymbol{x}$ with price related terms
  - Includes price and cross-terms with price

$$\boldsymbol{x} = [\boldsymbol{x}_c, p.\boldsymbol{x}_p]^T$$

Epistemic uncertainty, uncertainty in model parameters, captured via Bayesian framework: learn posterior distribution over $\boldsymbol{\beta}$

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Bayesian Logistic Model

$$\pi(\beta)$$

PRIOR

$$\sigma(p, \boldsymbol{x}|\boldsymbol{\beta}) = \frac{1}{1+e^{\boldsymbol{\beta}^T \boldsymbol{x}}}$$

$$y|\boldsymbol{x}, \boldsymbol{\beta} = \begin{cases} 1, w.p. & \sigma(\boldsymbol{x}|\boldsymbol{\beta}) \\ 0, w.p. & 1 - \sigma(\boldsymbol{x}|\boldsymbol{\beta}) \end{cases}$$

DATA LIKELIHOOD DISTRIBUTION

BAYESIAN INFERENCE

- Natural way to specify the uncertainty in the model parameters

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{x}) = \frac{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\beta})\pi(\boldsymbol{\beta})}{p(\boldsymbol{y})}$$

- Full distribution is available, can apply data efficient MAB heuristics
  - Thompson Sampling (TS)
  - Upper Confidence Bound (UCB)

CHALLENGES

- Posterior distribution hard to compute: doesn't belong to conjugate family of distributions
- High dimensional problem makes sampling-based methods like MCMC prohibitively expensive and slow

Use Variational Inference framework to construct fast and scalable algorithms

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

The Variational Bayes Framework

Bayesian Inference and Posteriors

$$\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) = \frac{p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\beta})\pi(\boldsymbol{\beta})}{p(\boldsymbol{y})}$$

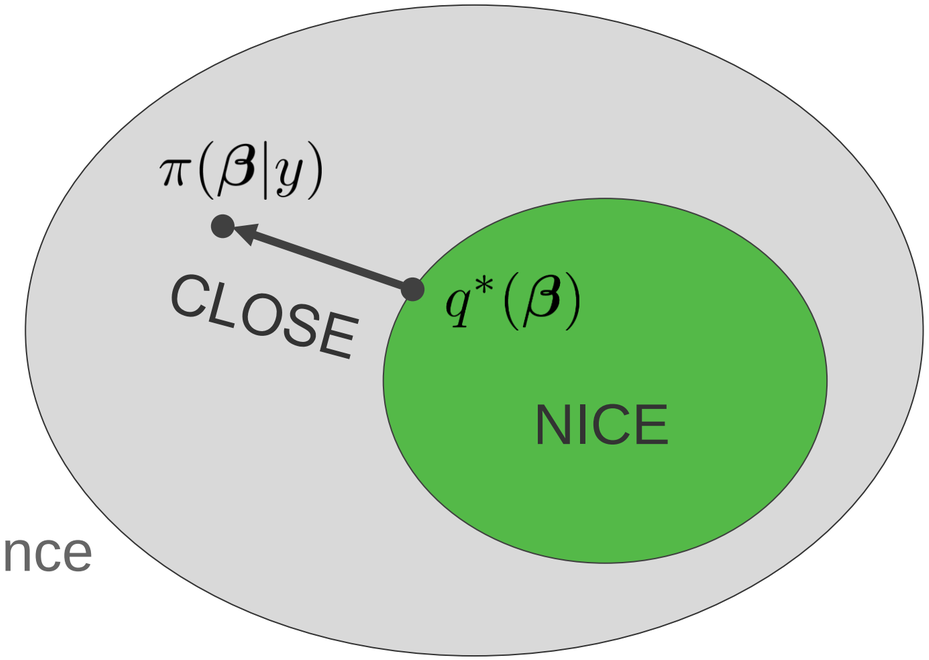
- Posterior $\pi(\boldsymbol{\beta}|y)$ is difficult to compute analytically

- Approximate posterior $\boldsymbol{q}^*(\boldsymbol{\beta})$ from a "nice" set of distributions by minimizing some notion of closeness

$$q^* = \arg\min_{q \in Q} f(q(\cdot), \pi(\cdot|y))$$

- **Variational Bayes**: $f(\cdot)$ is the Kullback-Leibler divergence

$$\eta^* = \arg\min_{\eta} KL\Big(q(\beta;\eta)||\pi(\beta|\boldsymbol{y},\boldsymbol{x})\Big)$$



Broderick, 2018 ICML

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Bayesian Logistic Model using Variational Bayes

$\pi(\beta)$

PRIOR

$$\sigma(\boldsymbol{x}|\boldsymbol{\beta}) = \frac{1}{1+e^{\boldsymbol{\beta}^T\boldsymbol{x}}}$$

$$y|\boldsymbol{x},\boldsymbol{\beta} = \begin{cases} 1, w.p. & \sigma(\boldsymbol{x}|\boldsymbol{\beta}) \\ 0, w.p. & 1-\sigma(\boldsymbol{x}|\boldsymbol{\beta}) \end{cases}$$

DATA DISTRIBUTION

- Prior distribution: Multi-variate Normal (Approximating distribution q)

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \leftarrow q(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\Sigma})$$

- Posterior density

$$\pi(\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{x}) \propto \frac{1}{1+e^{(1-2y)\boldsymbol{\beta}^T\boldsymbol{x}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}^T-\boldsymbol{\theta})}$$

- Posterior update via Variational Bayes

$$(\theta^*, \Sigma^*) = \arg\min_{\theta,\Sigma} KL_{\theta,\Sigma}(q||\pi)$$

Equivalently, maximize Evidence Lower Bound (ELBO) [Jordan et al., 1999; Bishop, 2006]

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Bayesian Logistic Model using Variational Bayes

- Our posterior update becomes an optimization problem, maximize ELBO

$$\max_{\theta,\Sigma} \mathbb{E}_{q(\beta|\theta,\Sigma)}\left[\ln p(y,\beta|\mathbf{x}) - \ln q(\beta|\theta,\Sigma)\right]$$

- Optimization over $d + \frac{d(d+1)}{2}$ variables where $|\boldsymbol{\beta}| = d$, by expanding the expectation and deriving gradients in closed form [Qu et al 2019]

$$\nabla_{\theta,\Sigma}\mathbb{E}_{q(\beta|\theta,\Sigma)}\left[\ln p(y,\beta|\mathbf{x}) - \ln q(\beta|\theta,\Sigma)\right]$$

- Automatic or Black-box stochastic variational inference using automatic differentiation packages like autograd and Stochastic Optimization [Ragnathan et al. 2013, Blei et al. 2018]

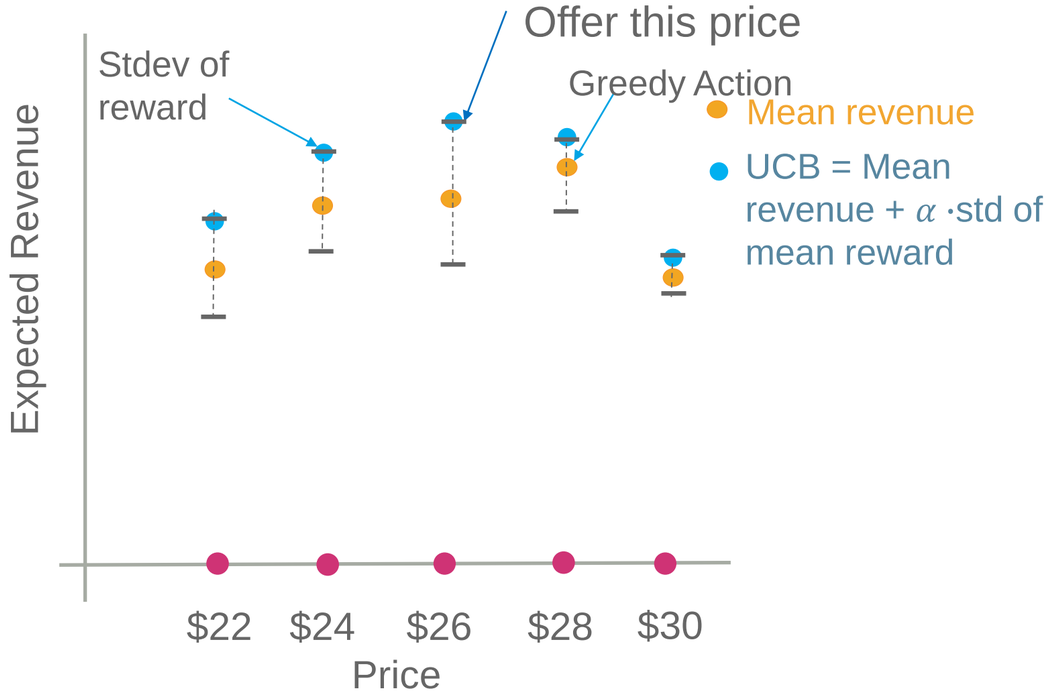
Fast updates: few seconds vs minutes in the case of MCMC

PROS. AutoGrad TensorFlow

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Multi-armed Bandit Heuristics with Bayesian Logistic Model

Upper Confidence Bound (UCB)



- $E[Rev(p, \boldsymbol{x}|\boldsymbol{\theta_t}, \boldsymbol{\Sigma_t})] \approx \frac{p}{1+e^{\boldsymbol{\theta_t}^T \boldsymbol{x}}}$
- $Std[Rev(p, \boldsymbol{x}|\boldsymbol{\theta_t}, \boldsymbol{\Sigma_t})] \approx \frac{pe^{\boldsymbol{\theta_t}^T \boldsymbol{x}}\sqrt{\boldsymbol{x}^T \boldsymbol{\Sigma_t} \boldsymbol{x}}}{\left(1+e^{\boldsymbol{\theta_t}^T \boldsymbol{x}}\right)^2}$

Thompson Sampling (TS)

$$\hat{\beta} \sim N(\boldsymbol{\theta_t}, \boldsymbol{\Sigma_t})$$

- $\hat{Rev}(p) \approx \frac{p}{1+e^{\hat{\boldsymbol{\beta}}_t^T \boldsymbol{x}}}$

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Simulation Test-Bed Tuned with Real Data

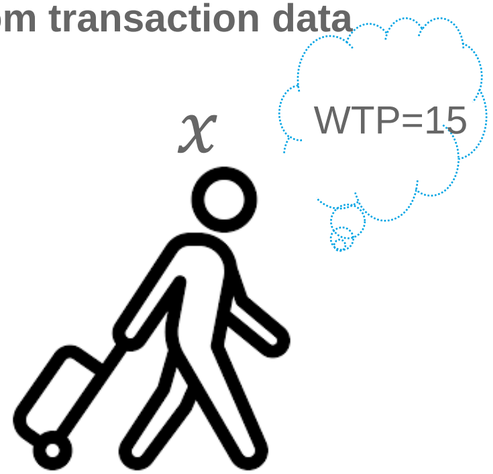
Consumer & Trip Profiles → WTP & Price of ancillary → Sample WTP and offer price → Determine consumer's decision

Generated by bootstrapping from transaction data

$\mu(x)$, $\sigma(x)$ **Tuned based on Real Data**

$$WTP \sim tN(\mu(x), \sigma(x))$$

x



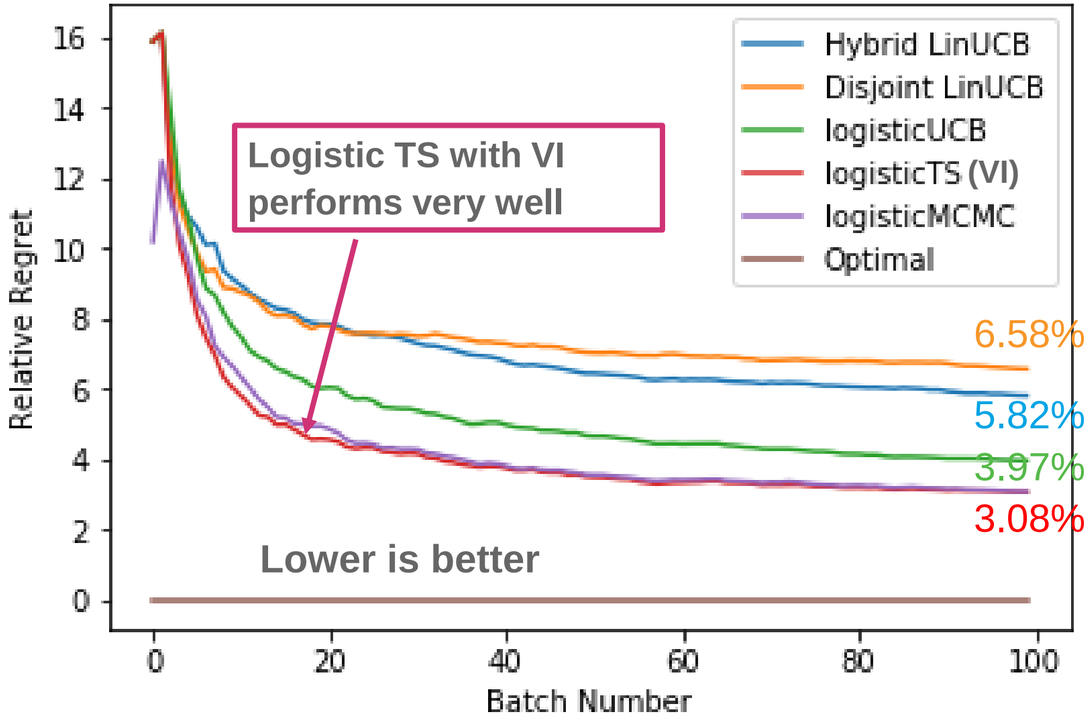
Buy as WTP > Offered Price

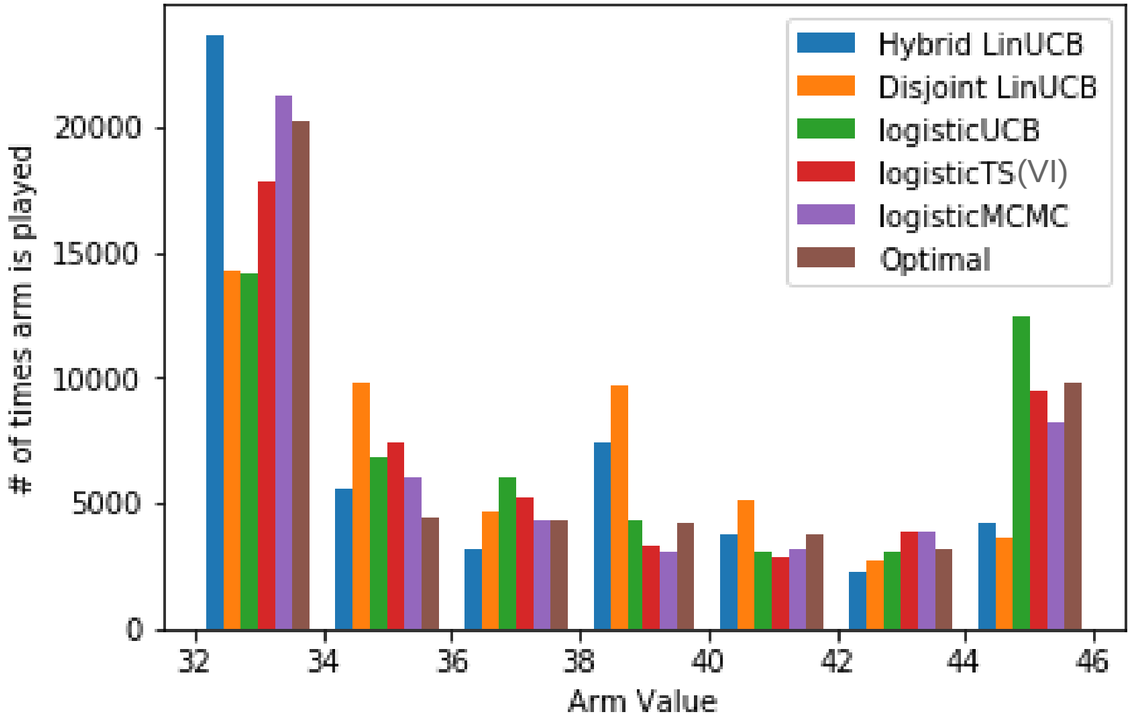
$14

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Simulation – Result

Base fare of Ancillary: $40 (Relative Regret = 15.2%)
Globally optimal price: $34 (Relative Regret = 9.1%)





$$\text{Relative Regret} = \frac{\text{Expected Cum. Optimal Revenue} - \text{Expected Cum. Revenue Under Policy}}{\text{Expected Cum. Optimal Revenue}} \times 100\%$$

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Conclusions

- Considered the problem of dynamic pricing when demand functions are unknown: specifically in the context of airline ancillary pricing
- Pricing problems have special structures: utilized Logistic MAB with Bayesian framework to develop data efficient heuristics
- Variational Bayes methodology can be used to construct fast and scalable algorithms for this setting
- Simulation experiments show that logistic bandit frameworks provide substantial benefits over widely used linear contextual bandits

Copyright © by Ravi Kumar/PROS. Published by AGIFORS with permission

Thank You

PROS Headquarters | Houston, Texas | 3100 Main Street, Suite 900 | Houston, TX 77002, USA | +1-713-335-5105 +1-800-555-3548