中国科学技术大学
UNI. OF SCI. & TECH. OF CHINA

Reviews and Self-Selection Bias with Operational Implications

Management Science，2021

Ningyuan Chen, Anran Li, Kalyan Talluri

中国科大·管理学院
SCH OF MNGT, UNI. OF SCI. & TECH. OF CHINA

CONTENTS

Authors

Ningyuan Chen(陈宁远)

- Associate Professor
  - Department of Management, University of Toronto Mississauga (Rotman School of Management, University of Toronto)
- Education
  - Ph.D. in Operations Research, **Columbia University**, 2015
  - M.S. in Operations Research, **Columbia University**, 2012
  - B.S. in Mathematics, **Peking University**, 2010
- Research Interests
  - Various approaches to making data-driven decisions in the context of revenue management
  - Statistics

Authors

Li, Anran（李安然）

- Associate Professor
  - Department of Decisions, Operations and Technology at The Chinese University of Hong Kong (CUHK) Business School
- Education
  - **Columbia University**, Ph.D., Operations Research, July 2017
    - Advisor: Guillermo Gallego
    - Dissertation Topic: “Product Line Design, Pricing and Framing under General Choice Models”
  - M.S., **Industrial Engineering**, December 2011
  - **Southwestern University of Finance and Economics**, B.S., Information Management and Information System July 2010
- Research Interests: Approximation Algorithm ; Dynamic Programming and Online Algorithm

Authors

Kalyan Talluri

- 2014-present Imperial College Business School, London, U.K, Professor of Operations Management
- Director, MSc in Business Analytics (on-campus and part-time)
- Fellow, Center for Business Analytics
- Fellow, Gandhi Center for Inclusive Innovation
- Education
  - **Massachusetts Institute of Technology**, Cambridge, MA, Ph.D. in Operations Research, 1991
    - Thesis Title: Issues in the Design and Analysis of Survivable Networks
  - **Purdue University**, West Lafayette, IN, MS in Industrial Engineering, 1987
  - **Osmania University**, Hyderabad, India, BE in Mechanical Engineering, 1985
- Research Interests: Business analytics; Revenue Management; Network and Service Design

Introduction

Background

- **Reviews and ratings reflects product qualities generally**

- ***acquisition bias***
  - Ratings may be a **positively skewed** indicator of the **true quality**
  - Consumers' **bounded rationality** can not separate **tastes and preferences**
    - *How to relate the magnitude of the bias to product attributes?*
    - *How is the market share affected by self-selection biases?*

Contribution

1、**Quantify** the acquisition bias and the impact on the rating of an arriving customer.

- The expected bias is **proportional** to the logarithm of the reciprocal of the **choice probability**.
- The model predicts that acquisition bias serves as the "**negative-reinforcing**" mechanism; it **offsets the intrinsic quality gap** between products.

2、Characterize the **asymptotic outcome** of **social learning**

- The average ratings and the choice probability of all products **converge to a limit**;
  - The limiting choice probability resembles the **multinomial logit (MNL)** model.
- Under social learning, the products look less **dissimilar quality-wise**;

3、Show how biases and social learning affect the **optimal assortment** and **pricing decisions**.

Related Literature

Table 1. Comparison of Model Setups in Li and Hitt (2008), Hu et al. (2017), Acemoglu et al. (2018), Besbes and Scarsini (2018), Vaccari et al. (2018), and Our Paper

	Li and Hitt (2008)	Hu et al. (2017)	Acemoglu et al. (2018)	Besbes and Scarsini (2018)	Vaccari et al. (2018)	Our paper
Acquisition bias	X	X	X	X		X
Underreporting bias		X				X
Bounded rationality	X	X		X	X	X
Rational Bayesian update			X	X		
Asymptotics			X	X	X	X
Multiple products					X	X
Firm's pricing and assortment decision						X

Model and Research Questions

Model Setup

- The firm is offering $d > 0$ products, q_i is true intrinsic quality of product $i, i \in \{1, \dots, d\}$, and p_i is price of product i, given exogenously.
- $\hat{q}_i(n)$: upon the arrival of customer n, the customer observes the average rating of all the products
- $\varepsilon_i(n)$: idiosyncratic preference realized before the purchase, taken to be i.i.d. mean-zero Gumbel random variable with CDF:

$$P(\varepsilon_i(n) \leq x) = e^{-e^{-(x-\mu)/\beta}}$$

- Ex ante net utility of owning product i:

$$u_i(n) = q_i - p_i + \varepsilon_i(n)$$

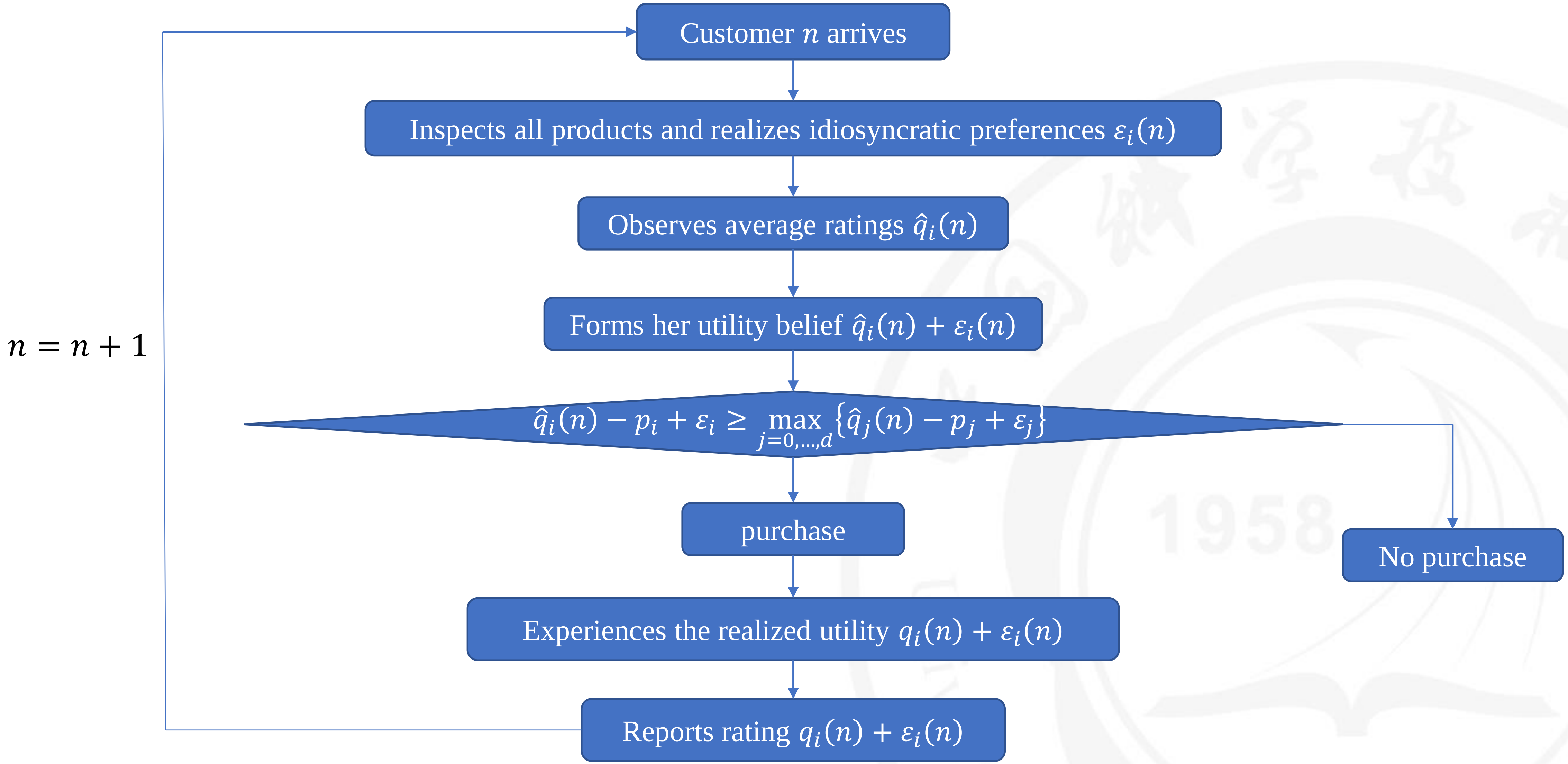
- The customer's choice probability is:

$$P\left(\hat{u}_i(n) \geq \max_{j=0,1,\dots,d} \hat{u}_j(n)\right) = \frac{\exp\left((\hat{q}_i(n) - p_i) / \beta\right)}{1 + \sum_{j=1}^{d} \exp\left((\hat{q}_j(n) - p_j) / \beta\right)}$$

- $\xi_i(n)$: experience shock $\xi_i(n) = q_i(n) - q_i$ is a mean-zero random variable, independent everything else
- $q_i(n)$: experienced quality of product i $q_i(n) \triangleq q_i + \xi_i(n)$

Model and Research Questions

Customer’s Purchasing and Rating Process



Model and Research Questions

Ratings Evolution Mechanism

- $N_i(n)$: the number of customers purchasing product i and reporting their ratings prior to customer n
- The rating system is fully characterized by:

$$X(n) \triangleq \left(\hat{q}_1(n), \dots, \hat{q}_d(n), N_1(n), \dots, N_d(n)\right)$$

- The average rating $q_i(n)$ is updated by the firm as follows:

$$\hat{q}_i(n+1) = \frac{N_i(n)}{N_i(n)+1}\hat{q}_i(n) + \frac{1}{N_i(n)+1}r_i(n)$$

- For all products $j \neq i$, their ratings remain the same:

$$\hat{q}_j(n+1) = \hat{q}_j(n)$$

- This paper does not specify the initial state of the ratings and start with $\hat{q}_i(0) \equiv 0$

Do Biases Exist?

- The event that customer n purchases product i is:

$$E_i(n) \triangleq \left\{\hat{u}_i(n) \geq \max_{j=0,1,\ldots,d} \hat{u}_j(n)\right\} = \left\{\hat{q}_i(n) - p_i + \epsilon_i(n) \geq \max_{j=0,1,\ldots,d} \hat{q}_j(n) - p_j + \epsilon_j(n)\right\}$$

- The expected rating is $E\{r_i(n)|E_i(n)\}$, so the **expected bias** is $E\{r_i(n) - q_i|E_i(n)\}$
- The ex post rating can be decomposed into $q_i + \varepsilon_i(n)$ and $\xi_i(n)$, the conditional distribution of the first component satisfies:

$$P(q_i + \varepsilon_i(n) \leq x|E_i(n)) = \frac{P\big(q_i + \varepsilon_i(n) \leq x, E_i(n)\big)}{P\big(E_i(n)\big)}$$

- The probability of $E_i(n)$ is:

$$C_i(n) \triangleq P\big(E_i(n)\big) = \frac{\exp\big((\hat{q}_i(n) - p_i)/\beta\big)}{1 + \sum_{j=1}^{d} \exp\big((\hat{q}_j(n) - p_j)/\beta\big)}$$

Do Biases Exist?

- The probability $P\big(q_i+\varepsilon_i(n)\le x, E_i(n)\big)$ is :

$$\underbrace{\int_{-\infty}^{x-q_i}\frac{1}{\beta}e^{(\mu-y)/\beta e^{-e(\mu-y)/\beta}}}_{\text{PDF of } \varepsilon_i(n)}\times\underbrace{\prod_{j=0,\dots,d,j\neq i}P\big(\hat{q}_j(n)-p_j+\epsilon_j(n)\le\hat{q}_i(n)-p_i+y\big)}_{P\left(\hat{q}_j(n)-p_j+\epsilon_j(n)\le\hat{q}_i(n)-p_i+y\right)\ \forall i\neq j}dy$$

- After simplifying $P(q_i+\varepsilon_i(n)\le x|E_i(n))$, we have

Proposition 1. Conditional on the purchase event $E_i(n), q_i+\varepsilon_i(n)$ has a Gumbel distribution with mean $q_i-\beta\log\big(C_i(n)\big)$ and variance $\pi^2\beta^2\ /\ 6$

$$P(q_i+\varepsilon_i(n)\le x|E_i(n)) = \exp\big(-e^{(\mu-(x-q_i+\beta\log(C_i(n)))/\beta)}\big) \quad (8)$$

Do Biases Exist?

- Therefore, conditional on a purchase of product i, the rating $r_i(n)$ of customer n satisfies:

$$E\{r_i(n)|E_i(n)\} = q_i - \beta \log\left(C_i(n)\right)$$

$$Var(r_i(n)|E_i(n)) = \frac{\pi^2\beta^2}{6} + \sigma_i^2$$

- The bias **always exists** and is **always positive**
- The customer reports a rating that is **higher** than the true quality in expectation
- The acquisition bias is **proportional** to the **logarithm of the reciprocal** of the choice probability

Model and Research Questions

Do the Average Ratings Converge? If so, to What?

- The limiting choice probability is given by

$$C_i^\infty = \frac{\exp\left((\hat{q}_i^\infty - p_i) / \beta\right)}{1 + \sum_{j=1}^d \exp\left((\hat{q}_j^\infty - p_j) / \beta\right)}$$

- The limiting bias must in turn be consistent with $\hat{q}_i^\infty$, i.e.,:

$$q_i - \beta \log(C_i^\infty) = \hat{q}_i^\infty$$

Theorem 1. As $n \to \infty$, $\hat{q}_i(n) \to \hat{q}_i^\infty$ and $C_i(n) \to C_i^\infty$ for all i almost surely, where

$$C_i^\infty = \frac{2e^{(q_i - p_i)/(2\beta)}}{\sum_{j=1}^d e^{(q_j-p_j)/(2\beta)} + \sqrt{\left(\sum_{j=1}^d e^{(q_j-p_j)/(2\beta)}\right)^2 + 4}}$$

$$\hat{q}_i^\infty = q_i - \beta \log(C_i^\infty)$$

Research Scene and assumption

- **Research Scene：** an online retailer selling products **with a rating system**

- **Research contents**:

  Explore how the social learning mechanism affects the firm’s optimal operational decisions in assortment planning and pricing.

- **Assumption:**
  - the seller is aware of the true qualities and the social learning mechanism,
  - customers behave according to the model of Section 3(base on the average rating)

- **The optimal assortment problem:** $$\max_{S\subseteq\{1,\ldots,d\}} \sum_{i\in S} p_i C_i^{\infty}(S)$$

  The optimal pricing problem: $$\max_{p_1,\ldots,p_d} \sum_{i=1}^{d} p_i C_i^{\infty}$$

Structural results of the optimal assortment

- Denote S^* as the optimal assortment under social learning:

Proposition 4. Suppose a product $i \in S^*$

- If $p_j e^{(q_j - p_j)/(2\beta)} > p_i e^{(q_i - p_i)/(2\beta)}$ *and* $e^{(q_j - p_j)/(2\beta)} < e^{(q_i - p_i)/(2\beta)}$, then $j \in S^*$
- If $p_j > 2p_i$, then $j \in S^*$.
- The product with the highest revenue is always in the optimal assortment.

Proposition 5. The assortment optimization problem (12) is NP-hard:

Proposition 6. The best revenue-ordered assortment can generate at least $1/2$ of the optimal revenue.

Monte Carlo Simulation

Table 2. Results of the Monte Carlo Simulation

d	*RConv200*	*RCon1000*	*RevMNL*	*RevSL*	*SurMNL*	*SurSL*	*RBias*	*RevDiff, %*	*RevOrder, %*
2	0.151	0.079	1.129	1.192	1.697	1.152	1.125	0.1	0.0
5	0.394	0.125	1.103	1.158	2.290	2.092	2.094	0.9	0.0
10	1.118	0.286	1.108	1.146	3.041	2.866	2.931	1.8	0.0
20	2.497	1.003	1.094	1.103	3.910	3.509	3.791	2.4	0.0

Results:

- The convergence to the limit may get slower when d increases as each product.
- The revenues under social learning are slightly higher.
- In practice, products of high quality are usually associated with a high price.
- The bias of social learning is significant, especially compared with the magnitude of the qualities ([0.5,2]).

Conclusion

Main Work and Results

- Quantify the acquisition bias when customers report ratings on a platform.
- The acquisition bias makes customer tastes appear more heterogeneous than they actually are.
- The acquisition bias benefits niche products and hurts popular products in terms of their market share.
- Point out the implications of this for the firm in its pricing and assortment optimization decisions

Future Research

- The links between reviews, quality, and the firm’s operational and managerial decisions.

Thanks for your attention!