Data-Driven Demand Learning and Dynamic Pricing Strategies in Competitive Markets

Demand Estimation

Rainer Schlosser, Martin Boissier, Matthias Uflacker

Hasso Plattner Institute (EPIC)

May 8, 2018

Outline

- Questions/Support: Market simulation (1st Exercise)
- Goals of today's meeting: Demand estimation
- How to estimate sales probabilities: Simple approaches
- 2nd Exercise: Demand learning using logistic regression

Customer Choice: Buying Books on Amazon

A Course in In-Memory Data Management: The Inner Mechanics of In-Memory Databases (Gebundene Ausgabe)

von Hasso Plattner (Autor)

Schreiben Sie die erste Bewertung

Optimieren durch Alles löschen

Versand
- [ ] Prime
- [ ] Versandkostenfrei

Zustand
- [ ] Neu
- [x] Gebraucht
  - [x] Wie neu
  - [x] Sehr gut
  - [x] Gut
  - [x] Akzeptabel

Preis + Versand (inkl. USt)	Zustand	Verkäufer-Information	Lieferung
EUR 44,90 + EUR 3,00 Versandkosten	**Gebraucht - Akzeptabel** Einband intakt und in sehr gutem Zustand, einige Seiten haben kle... » Weitere Informationen	**ialvamani** 100% positiv. (4 alle Bewertungen) Verkäuferinformationen, Impressum, AGB, Widerrufsrecht.	• **Ankunft zwischen** April 26 - Mai 2. • Versandtarife
EUR 45,00 + EUR 3,00 Versandkosten	**Gebraucht - Sehr gut** Versand aus Deutschland / We dispatch from Germany via Air Mail. ... » Weitere Informationen	**lange_und_springer_antiquariat** 98% positiv in den letzten 12 Monaten. (28.584 Bewertungen insgesamt) Verkäuferinformationen, Impressum, AGB, Widerrufsrecht.	• **Ankunft zwischen** April 27 - Mai 2. • Versand aus Deutschland • Versandtarife
EUR 65,60 + EUR 3,00 Versandkosten	**Gebraucht - Wie neu** New, Excellent customer service. Satisfaction guaranteed!!	**Totalbookstore** 89% positiv in den letzten 12 Monaten. (439 Bewertungen insgesamt) Verkäuferinformationen, Impressum, AGB, Widerrufsrecht.	• **Ankunft zwischen** Mai 3-20. • Versandtarife
EUR 79,56 + EUR 3,00 Versandkosten	**Gebraucht - Sehr gut** Publisher: Springer Date of Publication: 2014 Binding: hard... » Weitere Informationen	**Herb Tandree Philosophy Books** 90% positiv in den letzten 12 Monaten. (338	• **Ankunft zwischen** Mai 2-6. • Versand aus Vereinigtes Königreich • Versandtarife

Customer Behavior

seller	price	quality	rating	feedback	shipping
k	p_k	q_k	r_k	f_k	c_k
1	44.90	akzeptabel	100%	4	5 Tage
2	**45.00**	**sehr gut**	**98%**	**28,584**	**6 Tage**
3	65.60	wie neu	89%	439	11 Tage
4	79.56	sehr gut	90%	338	10 Tage
. . .					
K			. . .		

A Seller's Perspective: Observable Data

period	**sale**	**price**	rank	competitor's prices	for product i (ISBN)			
t	$y_t^{(i)}$	$a_t^{(i)}$	$r_t^{(i)}$	$p_{t,1}^{(i)}$	$p_{t,2}^{(i)}$	$p_{t,3}^{(i)}$	$p_{t,4}^{(i)}$	… $p_{t,K}^{(i)}$
1	**0**	**19**	3	13	17	20	25	
2	**0**	**15**	2	13	17	20	25	
3	**1**	**10**	1	13	15	20	/	
4	**0**	**10**	1	13	15	20	22	
5	**1**	**12**	2	11	15	20	24	
6	**0**	**15**	3	11	14	20	24	
…								

Goal

- We have: Market data + Sales data
- We want: Optimize prices + Maximize expected profits
- We need: Sales probabilities for our offer prices
- We use: Regression models, e.g, Logistic regression

Approach: Maximum Likelihood Estimation

- Idea: (1) Choose a model + (2) Find the best calibration
- Example: Coin Toss
- Data: 01011101010001000101001000110000
- Model: Bernoulli Experiment with success probability p
- Calibration: Which model, i.e., which p explains our data best?

Our Model: Bernoulli Distribution

- Random variable Y sale occurred (1 yes, 0 no)
- Success probability $P(Y=1)=p$ and $P(Y=0)=1-p$
- Bernoulli distribution $P(Y=k)=p^k \cdot (1-p)^{1-k}, \quad k=0,1$
- (Binomial distribution) $P(Y=k)=\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \quad k=0,\ldots,n$ (n=1)

Likelihood Function

- Bernoulli distribution $P(Y=k) = p^k \cdot (1-p)^{1-k}\,, \quad k = 0,1$

- Consider observed data $\vec{y} = (y_1, \ldots, y_N)\,, \quad y_i \in \{0,1\}\,, \quad i = 1, \ldots, N$

- Probability for our data $P(Y_i = y_i) = p^{y_i} \cdot (1-p)^{1-y_i}\,, \quad y_i \in \{0,1\}$

- *Joint probability*

  (Likelihood Function)

$$P(Y_1 = y_1, \ldots, Y_N = y_N) = \prod_{i=1}^{N} P(Y_i = y_i)$$
$$= \prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i}$$

- Now, maximize the joint probability over the success probability p!

Maximize the Likelihood Function

- $\max P(Y_1 = y_1, \ldots, Y_N = y_N)$ i.i.d.: independent, identically distributed
- $\max\limits_{p} \prod_{i=1}^{N} P(Y_i = y_i)$
- $\max\limits_{p \in [0,1]} \prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i}$

Actually, we wanted to find the best *p*.

- $\arg\max\limits_{p \in [0,1]} \prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i}$

We are interested in First Order Conditions. Hence, we do not like products!

Monotone Increasing Transformations

- $$\arg\max_{p\in[0,1]}\left\{\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right\}$$

$$=\arg\max_{p\in[0,1]}\left\{5\cdot\left(\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right)+17\right\}$$? (linear)

$$=\arg\max_{p\in[0,1]}\left\{\left(\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right)^2\right\}$$?? (convex)

$$=\arg\max_{p\in[0,1]}\left\{\ln\left(\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right)\right\}$$??? (concave)

Log-Likelihood Function

$$\arg\max_{p} P(Y_1 = y_1, \ldots, Y_N = y_N)$$

$$= \arg\max_{p\in[0,1]} \left\{ \ln\left(\prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i} \right) \right\}$$

$$= \arg\max_{p\in[0,1]} \left\{ \sum_{i=1}^{N} \ln\left(p^{y_i} \cdot (1-p)^{1-y_i} \right) \right\}$$

$$= \arg\max_{p\in[0,1]} \left\{ \sum_{i=1}^{N} \left(\ln\left(p^{y_i} \right) + \ln\left((1-p)^{1-y_i} \right) \right) \right\}$$

$$= \arg\max_{p\in[0,1]} \left\{ \sum_{i=1}^{N} \left(y_i \cdot \ln(p) + (1-y_i) \cdot \ln(1-p) \right) \right\}$$

Optimization

- FOC: $$\frac{\partial}{\partial p} P(Y_1 = y_1, \ldots, Y_N = y_N) \overset{!}{=} 0$$

$$\sum_{i=1}^{N} \left(y_i \cdot \ln(p)' + (1 - y_i) \cdot \ln(1 - p)' \right) \overset{!}{=} 0$$

- Solve for *p*.

- 1 Variable, 1 Equation (Unique solution p^*)

- **Result**: Our data fits to the model $P(Y = 1) = p^*$ and $P(Y = 0) = 1 - p^*$.

Generalization: Demand Estimation On Amazon

- Regular price adjustments (e.g., time intervals of ca. 2 hours)

- Observation of market conditions (at the time of price adjustments)

  e.g., Competitors' prices, quality, ratings, shipping time, etc.

- Sales observations: Points in time

- Rare events, i.e., 0 or 1 sales between price adjustments (2 hours)

A Seller's Data Set

period	sale	price	rank	competitor's prices	for product i (ISBN)			
t	$y_t^{(i)}$	$a_t^{(i)}$	$r_t^{(i)}$	$p_{t,1}^{(i)}$	$p_{t,2}^{(i)}$	$p_{t,3}^{(i)}$	$p_{t,4}^{(i)}$	… $p_{t,K}^{(i)}$
1	**0**	**19**	3	13	17	20	25	
2	**0**	**15**	2	13	17	20	25	
3	**1**	**10**	1	13	15	20	/	
4	**0**	**10**	1	13	15	20	22	
5	**1**	**12**	2	11	15	20	24	
6	**0**	**15**	3	11	14	20	24	
…								

Estimation of Sales Probabilities

- Goal: Quantify sales probabilities as function of our offer price
- Idea: Sales probabilities should depend on market conditions
- Approach: Maximum Likelihood

(1) Choose family of models: Logistic function

(2) Define explanatory variables (based on our data)

(3) Calibrate model: Find model coefficients

(4) Result: Quantify sales probabilities for any market situation!

Explanatory Variables

- Data: Market situation in t: $\vec{s} = (t, p_1, ..., p_K, q_1, ..., p_K, r_1, ..., r_K, f_1, ..., f_K, ...)$

- Define explanatory variables (What could affect decisions?):

$x_1(a,\vec{s}) := 1$	(Intercept)
$x_2(a,\vec{s}) := \textit{price rank}$	(Rank of offer price within competitors' prices)
$x_3(a,\vec{s}) := a - \min_{k=1,...,K} p_k$	(Price difference to best competitor)
$x_4(a,\vec{s}) := \textit{quality rank}$	(Rank of our product condition)
$x_5(a,\vec{s}) := \#\textit{commercials}$	(Number of competitors with feedback >10000)
$x_6(a,\vec{s}) := \textit{combinations}$	(Number of comp. with better price + better quality)
$x_7(a,\vec{s}) := 1_{\{a \cdot 100 \bmod 10 = 9\}}$	(Psychological Prices)

. . .

One Family of Models: Logistic Function

- $P\left(Y=1 \mid \vec{x}(a,\vec{s})\right) := e^{\vec{x}'\vec{\beta}} / (1+e^{\vec{x}'\vec{\beta}})$

$$= \frac{\exp\left(\beta_1 \cdot x_1(a,\vec{s}) + \beta_2 \cdot x_2(a,\vec{s}) + \ldots\right)}{1+\exp\left(\beta_1 \cdot x_1(a,\vec{s}) + \beta_2 \cdot x_2(a,\vec{s}) + \ldots\right)} \in (0,1)$$

- There are other families, but this is a good family

- Maximum Likelihood Estimation:

  Find best $\vec{\beta}$ coefficients for our data $y_t, \vec{x}(a_t, \vec{s}_t)$, $t=1,\ldots,N$

Maximize the Log-Likelihood Function

- Recall:

$$\arg\max_{p} P(Y_1 = y_1, \ldots, Y_N = y_N) = \arg\max_{p \in [0,1]} \left\{ \sum_{i=1}^{N} \big(y_i \cdot \ln(p) + (1 - y_i) \cdot \ln(1 - p) \big) \right\}$$

- Now, we have the conditional probabilities:

$$\arg\max_{\vec{\beta}} P\big(Y_1 = y_1 \mid a_1, \vec{s}_1, \ \ldots \ , Y_N = y_N \mid a_N, \vec{s}_N\big)$$

$$= \arg\max_{\beta_m \in \mathbb{R}, m=1,\ldots,M} \left\{ \sum_{i=1}^{N} \left(y_i \cdot \ln\left(\frac{e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}}{1 + e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}} \right) + (1 - y_i) \cdot \ln\left(1 - \frac{e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}}{1 + e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}} \right) \right) \right\}$$

Optimization

- FOC: $$\frac{\partial}{\partial \vec{\beta}} P(Y_1 = y_1 \mid a_1, \vec{s}_1,\ \ldots\ , Y_N = y_N \mid a_N, \vec{s}_N) \overset{!}{=} 0$$

$$\sum_{i=1}^{N} \left(y_i \cdot \frac{\partial}{\partial \beta_m} \ln\left(\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}} \right) + (1-y_i) \cdot \frac{\partial}{\partial \beta_m} \ln\left(1 - \frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}} \right) \right) \overset{!}{=} 0,\quad m = 1,\ldots,M$$

- Solve **the system** for coefficients $\vec{\beta} = (\beta_1, \ldots, \beta_M)$
- M Variables, M Equations (Unique solution $\vec{\beta}^* = (\beta_1^*, \ldots, \beta_M^*)$)
- Result: Our data fits to the model $P\left(Y = 1 \mid \vec{x}(a,\vec{s})\right) := e^{\vec{x}(a,\vec{s})'\vec{\beta}^*} / (1 + e^{\vec{x}(a,\vec{s})'\vec{\beta}^*})$

Application of the Model Obtained

- Observe current market situation for a product: $\vec{s}$

- For any admissible offer prices a we can evaluate $\vec{x}(a,\vec{s})$ and obtain

$$P\left(Y=1 \mid \vec{x}(a,\vec{s})\right) := \frac{e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}{1+e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}$$

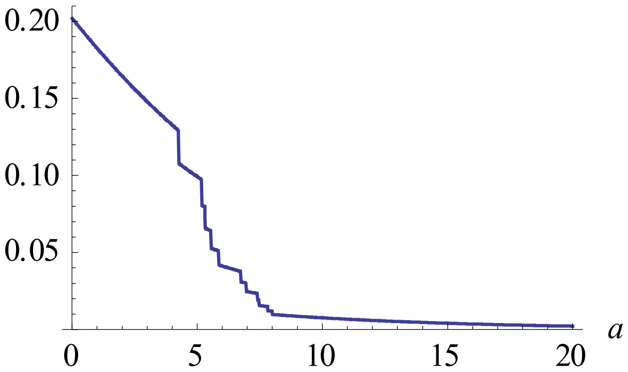
- Now, we can optimize expected profits (e.g., for one time interval):

$$\max_{a\geq 0}\left\{(a-c)\cdot\frac{e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}{1+e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}\right\}$$

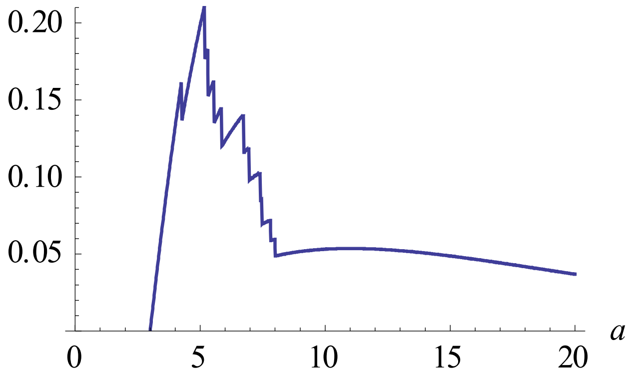
Prediction of Sales Probabilities

- Example: Competitor's prices $\vec{p} = (4.26,\ 5.18,\ 5.31,\ 5.55,\ 5.86,\ \ldots)$

$sales\ \ probability\ P(a)$



$expected\ \ profit\ \ (a-3)\cdot P(a)$



price

Summary

(+) Logistic Regression is simple and robust

(+) Allows for many observations N and many features M

(+) Plausibility Checks & Closed Form Expressions

(+/–) Definition of Customized Explanatory Variables

(–) No dependencies between variables

(–) Limited to binary dependent variables

What is a good Model?

- "Goodness of fit" measures (LL, AIC, McFadden Pseudo $R^2 := 1 - LL/LL0$)

- Logit: AIC (low is good, trade-of between fit and number of variables M)

$$AIC = -2 \cdot \sum_{i=1}^{N} \left(y_i \cdot \ln p_i + (1 - y_i) \cdot \ln(1 - p_i) \right) + 2 \cdot M = -2 \cdot LL + 2 \cdot M$$

Note, p_i depends on all features x_i and the optimal β^* coefficients.

- Be creative: Test different variables and find the smallest AIC value.

Hint: Not quantity but quality counts!

2nd Exercise – Demand Estimation

- Create random market situations with multiple sellers
- Choose a specific buying behavior (e.g., Scoring, Rank Based)
- Simulate sales events for different market situation
- Gather observable data and **estimate** sales probabilities (via logit model)
- Use different combinations of explanatory variables
- Compare the goodness of fit of different models

Overview

2	April 24	Customer Behavior
3	April 30	Pricing Strategies & DP, 1st Homework (market simulation)
4	May 8	Demand Estimation, 2nd Homework (demand learning)
5	**May 15**	**Introduction Price Wars Platform**
6	May 22	Warm up Platform Exercise (in Groups)
7	May 29	Dynamic Pricing Challenge / Projects
8	June 5	no Meeting
9	June 12	Workshop / Group Meetings
10	June 19	Presentations (First Results)
11	June 26	Workshop / Group Meetings
12	July 3	Workshop / Group Meetings
13	**July 10**	**Presentations (Final Results), Feedback, Documentation (Aug/Sep)**
14	July 17	no Meeting