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 (1\textsuperscript{st} Exercise)
- Goals of today’s meeting: Demand estimation
- How to estimate sales probabilities: Simple approaches
- 2\textsuperscript{nd} 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)

**Autor:** Hasso Plattner

<table>
<thead>
<tr>
<th>Preis + Versand (inkl. US$I)</th>
<th>Zustand</th>
<th>Verkäufer-Information</th>
<th>Lieferung</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EUR 44,90</strong></td>
<td><strong>Gebraucht - Akzeptabel</strong></td>
<td><strong>ialvamani</strong> ▫ ▫ ▫ ▫ 100% positiv, (4 alle Bewertungen)</td>
<td>▫ Ankunft zwischen April 26 - Mai 2.</td>
</tr>
<tr>
<td>+ EUR 3,00 Versandkosten</td>
<td></td>
<td>Verkäuferinformationen, Impressum, AGB, Widerrufsrecht</td>
<td>▫ Versandtanfe</td>
</tr>
<tr>
<td><strong>EUR 45,00</strong></td>
<td><strong>Gebraucht - Sehr gut</strong></td>
<td><strong>lange_und_springer_antiquariat</strong> ▫ ▫ ▫ ▫ 98% positiv in den letzten 12 Monaten. 28,564 Bewertungen insgesamt</td>
<td>▫ Ankunft zwischen April 27 - Mai 2.</td>
</tr>
<tr>
<td>+ EUR 3,00 Versandkosten</td>
<td></td>
<td>Verkäuferinformationen, Impressum, AGB, Widerrufsrecht</td>
<td>▫ Versand aus Deutschland</td>
</tr>
<tr>
<td><strong>EUR 65,60</strong></td>
<td><strong>Gebraucht - Wie neu</strong></td>
<td><strong>Totalbookstore</strong> ▫ ▫ ▫ ▫ 88% positiv in den letzten 12 Monaten. (438 Bewertungen insgesamt)</td>
<td>▫ Ankunft zwischen Mai 3-20.</td>
</tr>
<tr>
<td>+ EUR 3,00 Versandkosten</td>
<td></td>
<td>Verkäuferinformationen, Impressum, AGB, Widerrufsrecht</td>
<td>▫ Versandtanfe</td>
</tr>
<tr>
<td><strong>EUR 79,56</strong></td>
<td><strong>Gebraucht - Sehr gut</strong></td>
<td><strong>Herb Tandree Philosophy Books</strong> ▫ ▫ ▫ ▫ 90% positiv in den letzten 12 Monaten. 2008</td>
<td>▫ Ankunft zwischen Mai 2-6.</td>
</tr>
<tr>
<td>+ EUR 3,00 Versandkosten</td>
<td></td>
<td>Verkäuferinformationen, Impressum, AGB, Widerrufsrecht</td>
<td>▫ Versand aus Vereinigtes Königreich</td>
</tr>
</tbody>
</table>

*Publisher: Springer*<br>*Date of Publication: 2014*<br>*Binding: hard...*
### Customer Behavior

<table>
<thead>
<tr>
<th>seller ( k )</th>
<th>price ( p_k )</th>
<th>quality</th>
<th>rating</th>
<th>feedback ( f_k )</th>
<th>shipping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>44.90</td>
<td>akzeptabel</td>
<td>100%</td>
<td>4</td>
<td>5 Tage</td>
</tr>
<tr>
<td>2</td>
<td>45.00</td>
<td>sehr gut</td>
<td>98%</td>
<td>28,584</td>
<td>6 Tage</td>
</tr>
<tr>
<td>3</td>
<td>65.60</td>
<td>wie neu</td>
<td>89%</td>
<td>439</td>
<td>11 Tage</td>
</tr>
<tr>
<td>4</td>
<td>79.56</td>
<td>sehr gut</td>
<td>90%</td>
<td>338</td>
<td>10 Tage</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( K )</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
## A Seller’s Perspective: Observable Data

<table>
<thead>
<tr>
<th>period</th>
<th>sale</th>
<th>price</th>
<th>rank</th>
<th>competitor’s prices for product i (ISBN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$y_t^{(i)}$</td>
<td>$a_t^{(i)}$</td>
<td>$r_t^{(i)}$</td>
<td>$p_{t,1}^{(i)}$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>19</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>15</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
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: 010111010100010001010010001100000

- 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) $P(Y = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, \quad k = 0, \ldots, n \quad (n=1)$
Likelihood Function

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

- **Consider observed data** \( \tilde{y} = (y_1, ..., y_N) , \quad y_i \in \{0,1\} , \quad i = 1, ..., N \)

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

(Likelihood Function)

- **Now, maximize the joint probability over the success probability \( p \)!**
Maximize the Likelihood Function

- \( \max P(Y_1 = y_1, ..., Y_N = y_N) \) \quad \text{i.i.d.: independent, identically distributed}

- \( \max_p \prod_{i=1}^{N} P(Y_i = y_i) \)

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

Actually, we wanted to find the best \( p \).

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

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

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

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

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

\[ = \arg\max_{p \in [0,1]} \left\{ \ln \left( \prod_{i=1}^{N} p^{y_i} \cdot (1 - p)^{1-y_i} \right) \right\} \quad \text{(concave)} \]
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:** \[ \frac{\partial}{\partial p} P(Y_1 = y_1, \ldots, Y_N = y_N) = 0 \]
  \[ \sum_{i=1}^{N} (y_i \cdot \ln(p)' + (1 - y_i) \cdot \ln(1-p)') = 0 \]

- Solve for \( p \).

- 1 Variable, 1 Equation \quad (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

<table>
<thead>
<tr>
<th>period</th>
<th>sale $y_t^{(i)}$</th>
<th>price $a_t^{(i)}$</th>
<th>rank $r_t^{(i)}$</th>
<th>competitor’s prices for product $i$ (ISBN)</th>
<th>$p_{t,1}^{(i)}$</th>
<th>$p_{t,2}^{(i)}$</th>
<th>$p_{t,3}^{(i)}$</th>
<th>$p_{t,4}^{(i)}$</th>
<th>... $p_{t,K}^{(i)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>19</td>
<td>3</td>
<td></td>
<td>13</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>15</td>
<td>2</td>
<td></td>
<td>13</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td></td>
<td>13</td>
<td>15</td>
<td>20</td>
<td>/</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td></td>
<td>13</td>
<td>15</td>
<td>20</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td></td>
<td>11</td>
<td>15</td>
<td>20</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td></td>
<td>11</td>
<td>14</td>
<td>20</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 $t$: $\bar{s} = (t, p_1, ..., p_K, q_1, ..., p_K, r_1, ..., r_K, f_1, ..., f_K, ...)$

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

  $x_1(a, \bar{s}) := 1$ \quad ( Intercept )

  $x_2(a, \bar{s}) := \text{price rank}$ \quad (Rank of offer price within competitors’ prices)

  $x_3(a, \bar{s}) := a - \min_{k=1,...,K} p_k$ \quad (Price difference to best competitor)

  $x_4(a, \bar{s}) := \text{quality rank}$ \quad (Rank of our product condition)

  $x_5(a, \bar{s}) := \# \text{commercials}$ \quad (Number of competitors with feedback >10000)

  $x_6(a, \bar{s}) := \text{combinations}$ \quad (Number of comp. with better price + better quality)

  $x_7(a, \bar{s}) := 1_{\{a \mod 10 = 9\}}$ \quad (Psychological Prices)

  $\ldots$
One Family of Models: Logistic Function

- \( P(Y = 1 | \tilde{x}(a, \tilde{s})) := e^{\tilde{x} \tilde{\beta}} / (1 + e^{\tilde{x} \tilde{\beta}}) \)

\[
= \frac{\exp(\beta_1 x_1(a, \tilde{s}) + \beta_2 x_2(a, \tilde{s}) + ...)}{1 + \exp(\beta_1 x_1(a, \tilde{s}) + \beta_2 x_2(a, \tilde{s}) + ...)} \in (0, 1)
\]

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

- Maximum Likelihood Estimation:

Find best \( \tilde{\beta} \) coefficients for our data \( y_t, x(a_t, \tilde{s}_t), t = 1, ..., N \)
Maximize the Log-Likelihood Function

- Recall:

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

- Now, we have the conditional probabilities:

\[
\arg \max_{\beta} P \left( Y_1 = y_1 \mid a_1, \tilde{s}_1, \ldots, Y_N = y_N \mid a_N, \tilde{s}_N \right) = \arg \max_{\beta_m \in \mathbb{R}, m=1,...,M} \left\{ \sum_{i=1}^{N} \left( y_i \cdot \ln \left( \frac{e^{\tilde{x}(a_i, \tilde{s}_i)^{\prime} \beta}}{1 + e^{\tilde{x}(a_i, \tilde{s}_i)^{\prime} \beta}} \right) + (1 - y_i) \cdot \ln \left( 1 - \frac{e^{\tilde{x}(a_i, \tilde{s}_i)^{\prime} \beta}}{1 + e^{\tilde{x}(a_i, \tilde{s}_i)^{\prime} \beta}} \right) \right) \right\}
\]
Optimization

- **FOC:** 
  \[
  \frac{\partial}{\partial \beta} P(Y_1 = y_1 \mid a_1, \tilde{s}_1, \ldots, Y_N = y_N \mid a_N, \tilde{s}_N) = 0
  \]

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

- **Solve the system** for coefficients \( \tilde{\beta} = (\beta_1, \ldots, \beta_M) \)

- **M Variables, M Equations** \( \text{ (Unique solution } \tilde{\beta}^* = (\beta_1^*, \ldots, \beta_M^*) \text{ )} \)

- **Result:** Our data fits to the model 
  \[
  P(Y = 1 \mid \tilde{x}(a, \tilde{s})) := e^{\tilde{x}(a, \tilde{s})' \tilde{\beta}^*} / (1 + e^{\tilde{x}(a, \tilde{s})' \tilde{\beta}^*})
  \]
Application of the Model Obtained

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

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

$$P(Y = 1 | \tilde{x}(a,\tilde{s})) := \frac{e^{\tilde{x}(a,\tilde{s})'\beta^*}}{1 + e^{\tilde{x}(a,\tilde{s})'\beta^*}}$$

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

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

- Example: Competitor’s prices $\bar{p} = (4.26, 5.18, 5.31, 5.55, 5.86, \ldots )$
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-off 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 $\beta^*$ 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

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