Mathematical Modeling of Behavior

Motivation

Michel Bierlaire

Mathematical Modeling of Behavior

EPFL

Outline

Outline

Context

Simple example: model development

Simple example: quality of the estimates

Simple example: maximum likelihood estimation

Simple example: hypothesis testing

Simple example: forecasting

Motivation

Human dimension in

- engineering
- business
- marketing
- planning
- policy making

Motivation

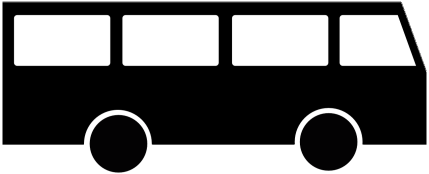
Concept of demand

Willingness and ability to purchase a commodity or service [Merriam-Webster]

Applications

Transportation

- Choice of destination
- Choice of transportation mode
- Choice of itinerary
- Choice of vehicle



Applications

Marketing

- Choice of packaging
- Choice of store
- Choice of product
- Choice of brand

Applications

Health

- Choice of treatment
- Choice of doctor
- Choice of training for doctors

Applications

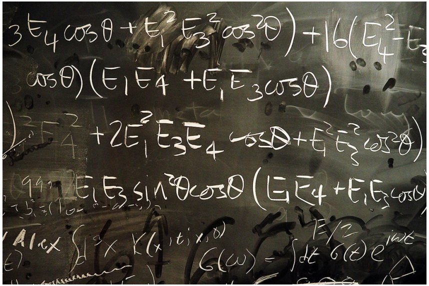
Energy

- Choice of appliances
- Choice of energy savings measures
- Choice of heating equipment

Motivation

Need for

- behavioral <u>theories</u>
- quantitative <u>methods</u>
- operational mathematical <u>models</u>



In this course...

Focus

- Individual / disaggregate behavior (vs. aggregate behavior)
- Theory of behavior which is
  - descriptive (how people behave) and not normative (how they should behave)
  - general: not too specific
  - operational: can be used in practice for forecasting
- Type of behavior: choice

Case studies

Mode choice in the Netherlands

- Context: car vs rail in Nijmegen.
- Objective: sensitivity to travel time and cost, inertia.

Mode choice in Switzerland

- Context: Car Postal.
- Objective: demand forecasting.

Case studies

Swissmetro

- Context: new transportation technology.
- Objective: demand pattern, pricing.

Residential telephone services

- Context: flat rate vs. measured.
- Objective: offer the most appropriate service.

Airline itinerary choice

- Context: questionnaire about itineraries across the US.
- Objective: help airlines and aircraft manufacturer to design a better offer.

Importance

Daniel L. McFadden

- UC Berkeley 1963, MIT 1977, UC Berkeley 1991
- Laureate of The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel 2000
- Owns a farm and vineyard in Napa Valley
- “Farm work clears the mind, and the vineyard is a great place to prove theorems”

Outline

Simple example

Objectives

Introduce basic components of choice modeling:

- definition of the problem
- data
- model specification
- parameter estimation
- model application

Application

Analysis of the market for electric cars

Choice problem

Choice

Consumer's choice to

- own an electric car
- own a car with combustion engine

Research questions

- what is the current market penetration of electric cars relative to combustion engine cars?
- how will the penetration change in the future?

Data

Population

- adults aged 20 and above
- in Switzerland
- owning a car

Sample

- 2500 individuals
- randomly selected

Questions

Is your car electric?

- Yes,
- No.

What is your age range?

- 20–39
- 40–64
- 65+

Data

Contingency table

	Age			
	20–39	40–64	65+	Total
Electric	65	55	5	125
Not electric	835	1045	495	2375
	900	1100	500	2500

Market penetration

- In the sample $125/2500 = 5\%$
- Currently in the population: by inference: 5%
- How do we predict? We need a model.

Model

Variables

- i: status of electric car ownership (yes or no)
- k: age category (20–39, 40–64 or 65$+$)

Model

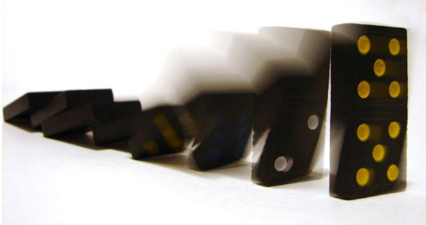
Decomposition

$$P(i,k) = P(i|k)P(k) = P(k|i)P(i)$$

Interpretation

- $P(i|k)$: age explains electric car ownership
- $P(k|i)$: electric car ownership explains age

Model



Model

- identify stable causal relationships between the variables
- stability over time necessary to forecast
- here: we select $P(i|k)$ as an acceptable behavioral model

Model

Specification

$$
\begin{aligned}
P(i = \text{yes} \mid k = \text{20–39}) &= \pi_1, \\
P(i = \text{yes} \mid k = \text{40–64}) &= \pi_2, \\
P(i = \text{yes} \mid k = \text{65+}) &= \pi_3.
\end{aligned}
$$

Parameters

- π_1, π_2, π_3
- unknown
- must be estimated from data

Model estimation

	Age			
	20–39	40–64	65+	Total
Electric	65	55	5	125
Not electric	835	1045	495	2375
	900	1100	500	2500

$$\pi_j = P(i = 1 | k = j) \approx \widehat{\pi}_j = \widehat{P}(i = 1 | k = j) = \frac{\widehat{P}(i = 1, k = j)}{\widehat{P}(k = j)}.$$

$$\pi_1 \approx \widehat{\pi}_1 = \widehat{P}(i = \text{yes} | k = \text{20–39}) = \frac{\widehat{P}(i = \text{yes}, k = \text{20–39})}{\widehat{P}(k = \text{20–39})} = \frac{65}{900} = 0.0722.$$

Model estimation

	Age			
	20–39	40–64	65+	Total
Electric	65	55	5	125
Not electric	835	1045	495	2375
	900	1100	500	2500

$$\pi_j = P(i = 1|k = j) \approx \widehat{\pi}_j = \widehat{P}(i = 1|k = j) = \frac{\widehat{P}(i = 1, k = j)}{\widehat{P}(k = j)}.$$

$$\pi_1 \approx \widehat{\pi}_1 = \widehat{P}(i = \text{yes}|k = \text{20–39}) = \frac{\widehat{P}(i = \text{yes}, k = \text{20–39})}{\widehat{P}(k = \text{20–39})} = \frac{65}{900} = 0.0722,$$

$$\pi_2 \approx \widehat{\pi}_2 = \widehat{P}(i = \text{yes}|k = \text{40–64}) = \frac{\widehat{P}(i = \text{yes}, k = \text{40–64})}{\widehat{P}(k = \text{40–64})} = \frac{55}{1100} = 0.0500,$$

$$\pi_3 \approx \widehat{\pi}_3 = \widehat{P}(i = \text{yes}|k = 65+) = \frac{\widehat{P}(i = \text{yes}, k = 65+)}{\widehat{P}(k = 65+)} = \frac{5}{500} = 0.0100.$$

Outline

Simple example: quality of the estimates

Motivation

- We have a model with parameters.
- We have used statistical inference to estimate the value of the parameters.
- This is subject to errors, as we have used the sample and not the population.
- How accurate are these estimates?

Informal checks

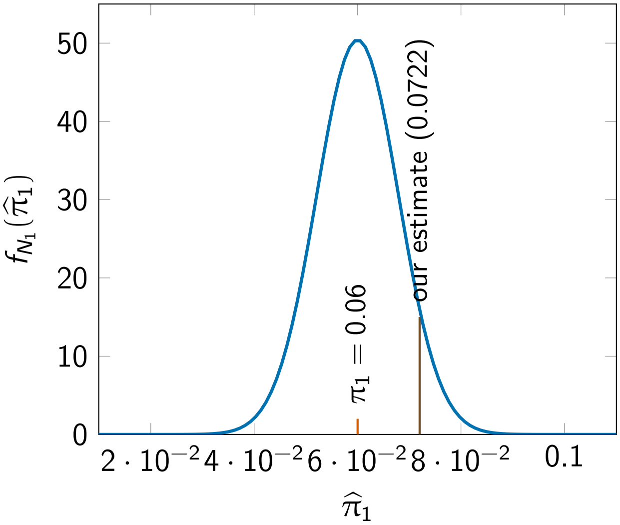
$$
\begin{aligned}
\widehat{\pi}_1 &= 65/900 = 0.0722, \\
\widehat{\pi}_2 &= 55/1100 = 0.0500, \\
\widehat{\pi}_3 &= 5/500 = 0.0100.
\end{aligned}
$$

- Do these estimates make sense?
- Do they match our a priori expectations?
- Here: as age increases, the market share of electric cars decreases.

Quality of the estimates

- How is $\widehat{\pi}_j$ different from π_j?
- We have no access to π_j.
- For each sample, we would obtain a different value of $\widehat{\pi}_j$.
- $\widehat{\pi}_j$ is distributed.

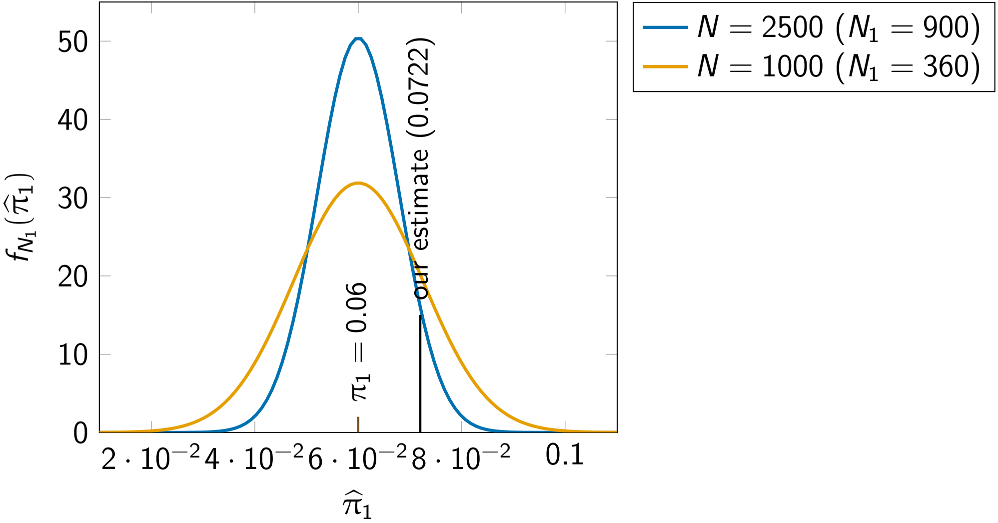
Distribution of $\widehat{\pi}_1$



Distribution of $\widehat{\pi}_1$

- Smaller samples are associated with wider spread.
- The larger the sample, the better the estimate.
- In practice, impossible to repeat the sampling multiple times.
- Distributions derived from theoretical results or simulation.

Distribution of $\widehat{\pi}_1$



Statistical properties

- Bernoulli (0/1) random variables.
- Variance: $\sigma_j^2 = \pi_j(1 - \pi_j)$.
- Sample average: unbiased estimator.
- Standard error of the estimator: $\sqrt{\sigma^2/N}$.
- Estimated standard error:

$$\widehat{s}_{\pi_j} = \sqrt{\widehat{\pi}_j(1 - \widehat{\pi}_j)/N_j}.$$

Outline

Simple example: maximum likelihood estimation

Motivation

- We investigate another estimator.
- Indeed, using the sample average is not possible for more complex models.
- Later, we will use the maximum likelihood estimator.
- We remind the concept on the simple example.

Maximum likelihood estimation

Likelihood

Probability that the model correctly predicts all the observations.

Likelihood function

$$\mathcal{L}^* = \prod_{n=1}^{N} P(i_n|k_n).$$

For our example

$$\mathcal{L}^* = \pi_1^{65}(1-\pi_1)^{835}\pi_2^{55}(1-\pi_2)^{1045}\pi_3^{5}(1-\pi_3)^{495}.$$

Maximum likelihood estimation

Estimates

- Values of the parameters that maximize $\mathcal{L}^*$.
- In practice, the logarithm is maximized

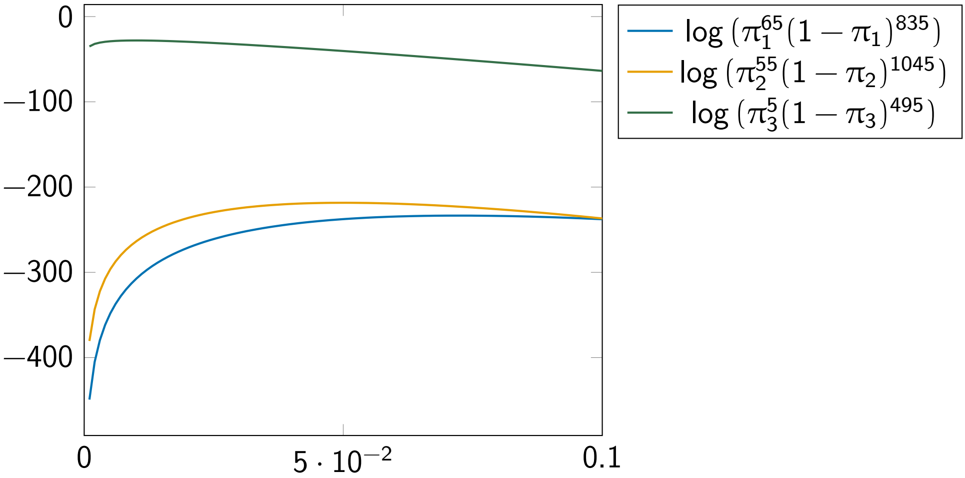
$$\mathcal{L} = \ln \mathcal{L}^* = \sum_{n=1}^{N} \ln P(i_n|k_n).$$

As $0 \leqslant \mathcal{L}^* \leqslant 1$, we have $\mathcal{L} \leqslant 0$.

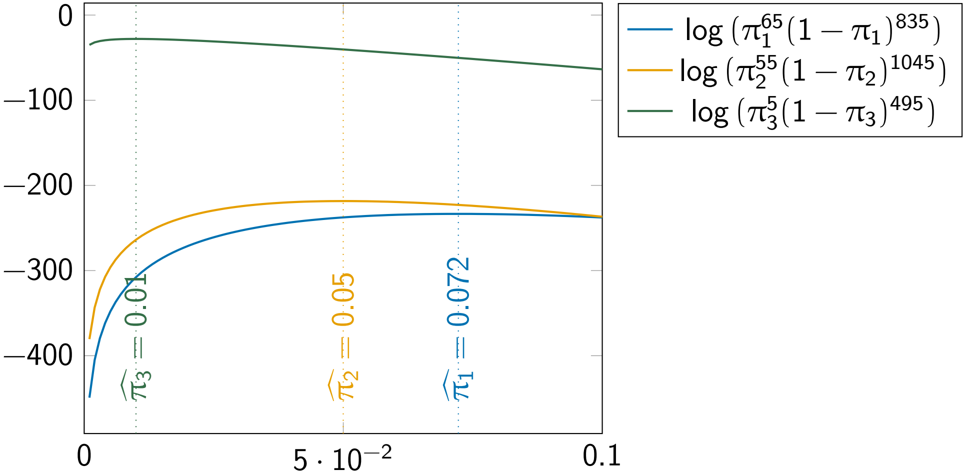
Properties

- Consistency.
- Asymptotic efficiency.

Maximum likelihood



Maximum likelihood



Maximum likelihood

Comments

- In this example, the maximum likelihood estimates happen to be the same as the sample average.
- For more complex models, the (log) likelihood function is not necessarily separable.
- The maximization is then more complicated to perform, but the concept is the same.

Outline

Hypothesis testing

Motivation

- Modeling assumption: age explains electric car ownership.
- We indeed observe that the three parameters are different.
- But this difference may be due to sampling errors.
- What can we say about our hypothesis?

Hypothesis testing

Null hypothesis

- Age does not explain electric car ownership.
- If it is true, then $\pi_1 = \pi_2 = \pi_3 = \pi$.
- But it does not mean that $\widehat{\pi}_1 = \widehat{\pi}_2 = \widehat{\pi}_3$.

Restricted model

$$\pi^{125}(1-\pi)^{2375}.$$

Unrestricted model

$$\pi_1^{65}(1-\pi_1)^{835}\pi_2^{55}(1-\pi_2)^{1045}\pi_3^{5}(1-\pi_3)^{495}$$

Likelihood ratio test

Intuition

If the null hypothesis is true, the estimates for the restricted and the unrestricted model should not be too different.

Formally

Under the null hypothesis, the statistic

$$-2(\mathcal{L}^R - \mathcal{L}^U)$$

is asymptotically distributed as χ^2 with degrees of freedom equal to the number of restrictions (in our case, 2).

Hypothesis testing

Unrestricted model

$$\widehat{\pi}_1 = 0.0722,\ \widehat{\pi}_2 = 0.0500,\ \widehat{\pi}_3 = 0.0100,\ \mathcal{L}^U = -479.782.$$

Restricted model

$$\widehat{\pi} = 0.05,\ \mathcal{L}^R = -496.288.$$

Statistic: likelihood ratio

$$-2(\mathcal{L}^R - \mathcal{L}^U) = 33.01$$

The probability to obtain such a value under the null hypothesis is lower than 10^{-5}. Therefore, we can safely reject the null hypothesis, and the unrestricted model is accepted.

Outline

Simple example: forecasting

Motivation

- At this stage, we have a tested model.
- We are ready to use it.
- We want to forecast the future market shares of electric cars.

Present situation

Age group	20–39	40–64	65+
Current share	36 %	44%	20%
Market penetration	7.2%	5%	1%

Total market penetration $= 36\%\ 7.2\% + 44\%\ 5\% + 20\%\ 1\% = 5\%$

Future scenario

Age structure will change in the future

Age group	20–39	40–64	65+
Current share	36 %	44%	20%
Future share	25 %	50%	25%
Market penetration	7.2%	5%	1%

Future total market penetration $= 25\%\ 7.2\% + 50\%\ 5\% + 25\%\ 1\% = 4.55\%$

Forecasting

- Causal relationship does not vary over time.
- Characterized by the model specification, including the values of its parameters.
- Values of the explanatory variables evolve over time.

Summary

- Motivation
- Simple example:
  1. definition of the problem,
  2. data collection,
  3. model specification,
  4. parameter estimation,
  5. hypothesis testing,
  6. model application.