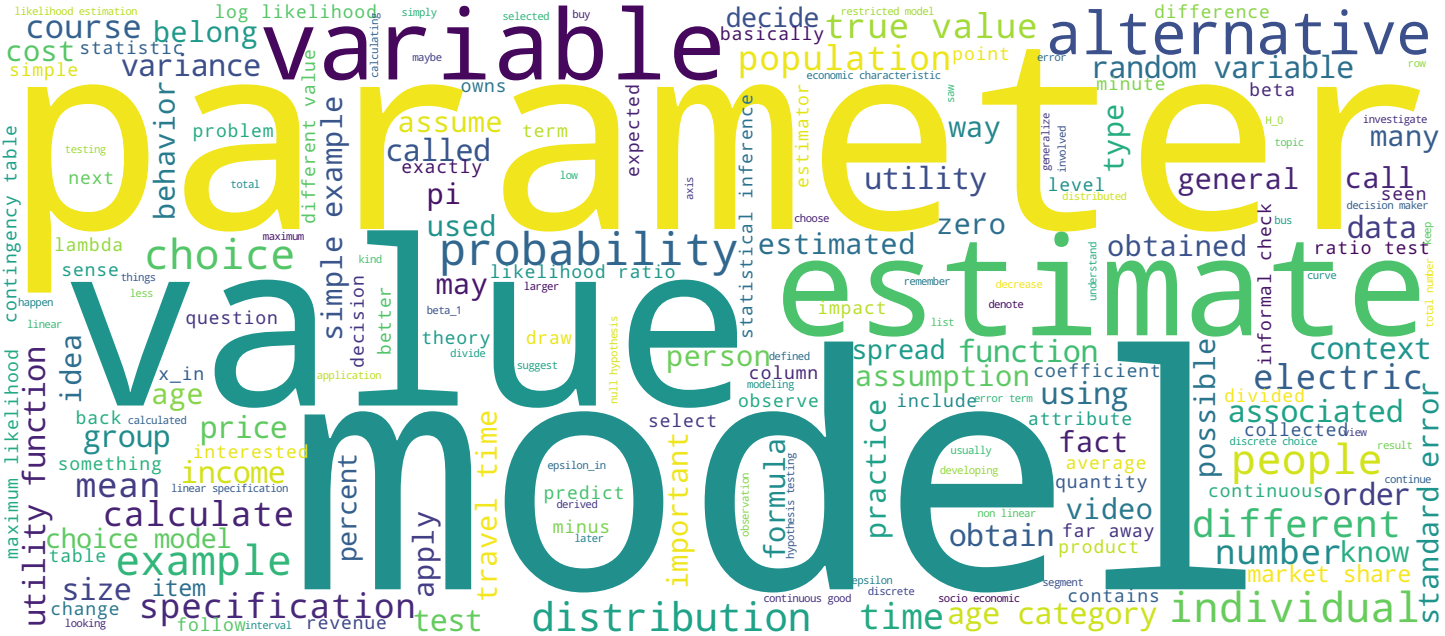
Introduction to behavior modeling

Simple example

Michel Bierlaire

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE


parameter
model
value
estimate
variable
alternative
probability
distribution
specification
example
choice
people
different
individual
time
age
category
income
price
calculate
mean
size
group
percent
apply
test
travel time
important
formula
practice
obtain
video
number
know
order
continuous
contains
socio economic
market share
standard error
utility function
maximum likelihood
contingency table
simple
behavior
variance
course
belong
cost
statistic
log likelihood
decide
basically
true value
population
difference
random variable
minute
beta
many
call
general
utility
level
type
way
zero
data
ratio test
obtained
estimated
statistical inference
estimator
expected
assume
called
exactly
pi
used
may
likelihood ratio
theory
draw
person
spread function
assumption
coefficient
using
attribute
fact
average
quantity
far away
product
associated
collected
discrete choice
divided
possible
informal check
context
electric
impact
observe
include
select
column
defined
modeling
error term
usually
developing
non linear
epsilon
discrete
continuous good
segment
contains
predict
derived
minus
later
epsilon_in
hypothesis testing
observation
x_in
age
back
calculated
price
group
interested
something
income
linear specification
mean
calculate
choice model
table
example
size
item
change
looking
specification
follow
interval
revenue
test
apply
percent

Simple example: quality of the estimates

We continue to investigate our simple example. We have been developing a model, we have specified it, it contains three parameters, pi_1, pi_2, pi_3, and using the data that we have collected, we have been able to estimate the value of these parameters. Clearly, these values have been calculated using statistical inference, assuming that what we calculate in the sample can generalize to the population. But we do mistakes by doing this. There are errors in these estimates. And can we have a sense of how accurate these estimates are? This is the topic of this video.

Notes

Summary

0m 04s

Parameters estimates

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

So these are the three values of the parameters that we have estimated from the contingency table. For example, pi hat 1 is 65 divided by 900. 65 is the number of people who own an electric car and belong to the first age category, and 900 is the total number of people who belong to age category 1. So, it is about 7.22%. And we do the same for the next two parameters.

Notes

Summary

0m 43s

Informal checks

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

In general, the first thing we do when we estimate parameters is to look at their values and do an informal check. Do they make sense? Their values, do they match our a priori expectations? Here, we observe that as age increases, the market share of electric car decreases. And actually, this is something that could be expected, because we expect younger people to be more attracted by this new technology. So this is not completely unexpected, and we may accept this kind of result. So that's the type of thing that we do first when we estimate the parameters of the model. Just informal checks.

Notes

Summary

1m 12s

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.

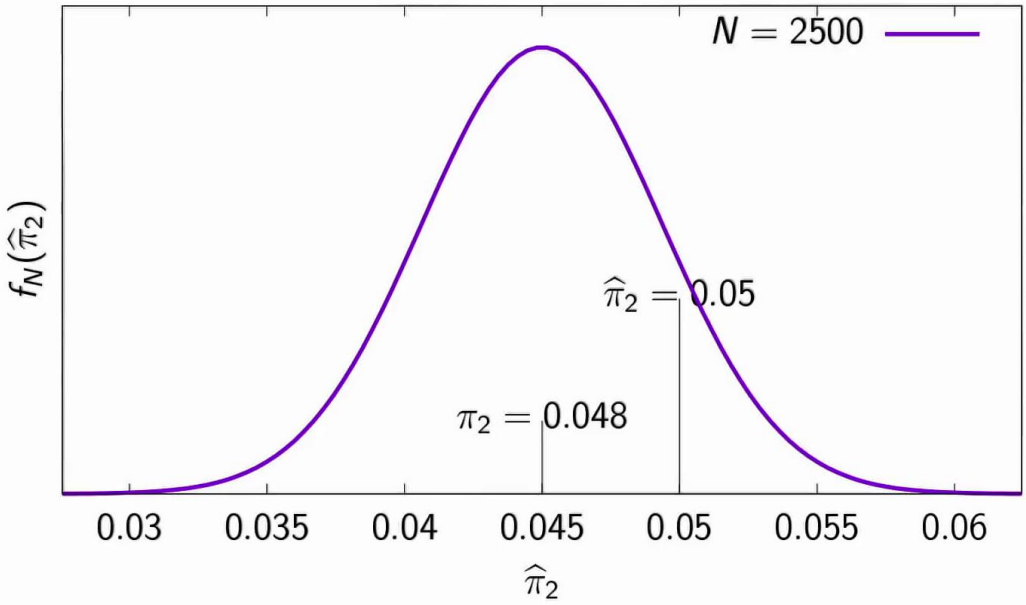
But this is not sufficient. We also want to do more formal and rigorous tests, to have an idea of the quality of the estimates. And because pi hat, the estimates of the parameters, have been obtained from statistical inference, the main question is: how is this value different from the true value of the parameter in the population. And, obviously, we have no access to the true value of the parameter from the population. This is why we do inference. Actually, each time you sample, you will obtain a different value for the estimate, for each different sample, because each sample is randomly drawn. So, therefore, this value of the estimate is distributed. Each time you draw a sample, you get a different value for pi hat j.

Notes

Summary

1m 56s

Distribution of $\widehat{\pi}_2$



So, for example, the distribution of pi hat two can be seen as this curve. In this example, I assume that the true value of pi_2 is equal to 0.048. But from the sample that we have collected, we have estimated that the value of the parameter is 0.05. It's different from the true value. And if I had drawn another sample, I would have found another value and all these values would follow this distribution. In general, all the estimators that we will be using are unbiased, or asymptotically unbiased. It means that the true value of the parameter is the expectation of the estimate. Of course, this distribution, we hope that it is not too spread. Because, if it is spread, we may end up with estimates which are far away from the true value of the parameter.

Notes

Summary

2m 41s

Distribution of $\widehat{\pi}_2$

- 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

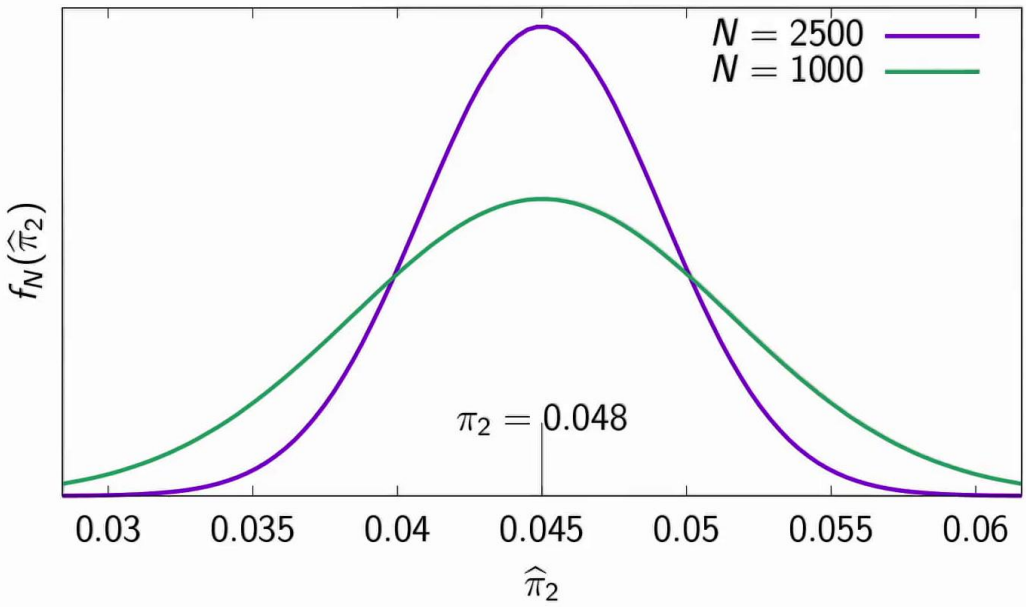
And the spread of the distribution is actually associated with the size of the sample. Smaller samples are associated with a wider spread. So, the larger the sample, the better the estimate. So it's always worth investing in large samples. But in practice, I cannot really analyze the distribution, the empirical distribution of the estimate. This would involve collecting several samples, and each time calculating the value of the estimates, and doing it again at again. In practice, it's not possible. Therefore, in general, the distributions of these estimators, are derived from theoretical results, or by simulation.

Notes

Summary

3m 35s

Distribution of $\widehat{\pi}_2$



If you change the sample size, the spread, that is called the variance of the distribution, increases. So in this example, I have plotted the original shape of the distribution of pi_2 hat in blue, for a sample size of 2500 that we used in our simple example. But if I calculate the distribution for a sample of size 1000, I obtain a distribution with a much bigger variance. And therefore, I have more chance to obtain an estimate that is far away from the true value of the parameter.

Notes

Summary

4m 13s

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:

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

In order to calculate statistical properties, we need to go back to the main random variables which are involved in this model. The random variables are very simple because each person can respond two things to the question: you can observe that he owns a car, he owns an electric car, or he does not own an electric car.So random variables which are 0 or 1 with some probability are called "Bernoulli" random variables. So, here, the probability that the random variable is 1 is pi_j, in age category j, and the variance of a Bernoulli variable is given by the probability of a success, times the probability of a failure. The probability that the person owns an electric car, times the probability that the person does not own an electric car. And, when you calculate the sample average, like we did in this example, we have an unbiased estimator of the true value of the parameter. And the standard error, basically the average spread around the mean, is calculated as the ratio between the variance and the sample size, and you take the square root of this. So the estimated standard error is obtained by calculating the variance using this formula (of course we don't know the true value of the parameter, so we replace it by the estimated value, pi hat j), and we divide by the number of people.

Notes

Summary

4m 51s

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:

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

So the standard error of the parameter associated with age category j, because it's calculated based on the number of people in this category, N_j is equal to the number of people in the age category j. So I suggest now that you calculate the standard errors using this formula of the estimates that you have obtained from the contingency table. And afterwards, we will look into a more complex estimator, which is called the maximum likelihood estimator, which is the one that we will use later in the course. Actually, in this case, in the case of the simple example, it will happen to be the same estimates as for the sample average. But we will look into this from the point of view of the maximum likelihood estimation.

Notes

Summary

6m 14s