Case Study 5 – Part 2: Bayesian Estimation

Prepared by Mazen Danaf, using code by Rico Krueger

Introduction

In this case study, we will use Bayesian estimation as an alternative to Maximum Likelihood which we have been using so far. We will be using the Swissmetro data set (which you are familiar with by now). We start by estimating Multinomial logit (MNL) and Logit Mixture (MXL) models using both estimation methods and comparing the estimates.

Afterwards, we use Hierarchical Bayes (HB) to estimate individual-specific parameters, and use these in individual predictions. We show that using individual-specific parameters achieves significantly better predictions, because they are conditional on the individual's previous choices. For this, you will need the following: "**mxl.py**", "**mxlMcmc.py**", and "**swissmetro_long.csv**"

Part 1: Multinomial Logit

In this section, we will estimate MNL and logit mixture models using both Biogeme and MCMC. We will consider the simple behavioral model shown below:

$$V_{CAR} = ASC_{CAR} + \beta_{time} Time_{CAR} + \beta_{Cost} Cost_{CAR}$$

$$V_{SM} = ASC_{SM} + \beta_{time} Time_{SM} + \beta_{Cost} Cost_{SM}$$

$$V_{TRAIN} = \beta_{time} Time_{TRAIN} + \beta_{Cost} Cost_{TRAIN}$$

We start by estimating the model in Biogeme using the code: "**SimpleMNL.py**". The results are shown below. The final log-likelihood value is 5331.

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	0.547	0.0461	11.9	0	0.049	11.2	0
ASC_SM	0.701	0.0549	12.8	0	0.0826	8.49	0
B_COST	-0.108	0.00518	-20.9	0	0.00682	-15.9	0
B_TIME	-0.128	0.00569	-22.5	0	0.0104	-12.3	0

We now compare our estimates to those obtained from MCMC estimation. This is done using the Metropolis Hastings (MH) algorithm. The code used in estimation is "Model1" in the workbook "**CS5Example**". The results are shown below. The final log-likelihood is also 5331.

	mean	std. dev.	2.50%	97.50%	R-hat
ASC_SM	0.698	0.056	0.586	0.806	1.011
ASC_CAR	0.545	0.046	0.458	0.634	1.016
B_COST	-1.083	0.048	-1.180	-0.990	1.015
B_TIME	-1.280	0.057	-1.396	-1.171	1.010

The R-hat values are close to 1, indicating that the Markov Chains are stationary for all parameters. Note that the posterior means are very close to the Biogeme estimates, and the posterior standard deviations are close to the Biogeme standard errors.

Part 2: Logit Mixture with Random Coefficients

Now we consider the case where the coefficients of time and cost are randomly distributed in the population. We use the same utility equations above, and start by assuming a normal distribution for the two parameters. Note that the normal distribution here is not ideal, because it is unbounded (*we expect these coefficients to be negative for everyone in the population*).

We first use Biogeme to estimate the model above using the code: "**RandomCoef.py**". The results with 5000 draws are shown below. The final log-likelihood of the model is 3920.88.

Name	Value	Std err	t-test	p-value	Rob. Std err	Rob. t-test	Rob. p-value
ASC_CAR	0.731	0.0807	9.06	0	0.14	5.21	1.94E-07
ASC_SM	0.358	0.0912	3.92	8.81E-05	0.147	2.43	0.0152
B_COST	-0.402	0.0309	-13	0	0.0379	-10.6	0
B_COST_S	-0.477	0.0345	-13.8	0	0.0429	-11.1	0
B_TIME	-0.473	0.0215	-21.9	0	0.0252	-18.8	0
B_TIME_S	0.43	0.019	22.6	0	0.0229	18.7	0

Next, we use the HB for logit mixture that we learned in class for logit mixture with random coefficients $\zeta_n \sim N(\mu, \Omega)$. This uses the same procedure described in your slides, and in Train (2009) (Chapter 12).

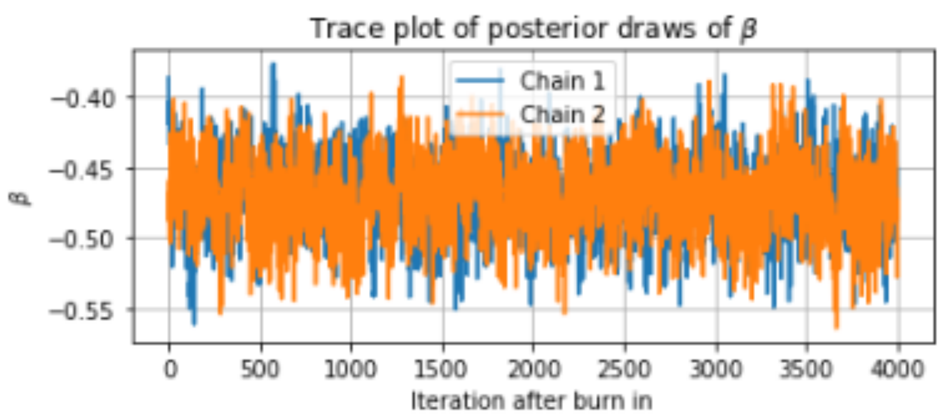
1. $\mu|\Omega, \zeta_n$: using a normal Bayesian update with unknown mean and known variance.
2. $\Omega|\mu, \zeta_n$: using a normal Bayesian update with known mean and unknown variance.
3. $\zeta_n|\mu, \Omega$: using a Metropolis-Hastings algorithm.

Note that the fixed parameters cannot be estimated using the same 3-Step Gibbs sampler. We will need an additional MH step to estimate these fixed parameters as in MNL. The results are shown below.

Fixed parameters:					
	mean	std. dev.	2.50%	97.50%	R-hat
ASC_SM	0.377	0.101	0.186	0.578	1.001
ASC_Car	0.749	0.084	0.583	0.916	1.002
Random parameters (means):					
	mean	std. dev.	2.50%	97.50%	R-hat
B_Cost	-0.411	0.029	0.356	0.468	1.003
B_Time	-0.470	0.026	-0.523	-0.420	1.004
Random parameters (standard deviations):					
	mean	std. dev.	2.50%	97.50%	R-hat
B_Cost	0.480	0.030	0.424	0.543	1.005
B_Time	0.443	0.022	0.403	0.488	1.009

The results are not numerically the same because we used simulation to estimate these models. Ideally, we want to use more draws. The final log-likelihood is -3920.1.

We can plot the Markov Chains of the estimated parameters to check if they are stationary. Let's consider the population mean of time (estimated as -0.47). The plot below shows two Markov Chains. By comparing the two, we can notice that they have the same mean and variance, and we can conclude they are stationary. This is also reflected in the value of R-hat, which is close to 1.



Part 3: Individual Predictions

Now we will use Bayesian estimation to obtain individual-specific parameters, that are the values of ζ_n conditional on the observed choices. We will see how these parameters improve predictions significantly, even on out-of-sample data. For this, we use the workbook "**Forecasting**". We start by dividing the Swissmetro data into training and test data sets. The training data set includes choices 1-8 for each individual, and the testing dataset is the 9^{th} (last) choice.

Here we are using the individual specific coefficients that we estimated for each survey respondent using his/her first 8 choices, in order to predict their 9^{th} choice.

The individual specific coefficients are the posterior estimates of ζ_n obtained from the Gibbs sampler. Unlike Maximum Likelihood, we do not need any post-processing in order to estimate these; they are already available (as we sampled from their conditional distribution in Step3 of estimation). The results are shown in the table below:

	MNL	MXL-Conditional	MXL-Unconditional
Likelihood	-649.12	-652.62	-481.2
Predicted Probability	0.497	0.511	0.694

The predicted probability of the chosen alternative on the test data is around 49.7% with MNL. If we estimate a logit mixture model and calculate the unconditional probability (without using the posteriors of the individual-specific coefficients), we achieve a slight improvement (51.1%). However, if we estimate the probability conditional on the previous choices, the predicted probability is 69.4%! This is how we can use logit mixture models in ***personalization and individual predictions.***

Your task:

1. Using the Swissmetro data, estimate a model where all four parameters are distributed (time, cost, and the two constants) on the first 8 choices.

2. Calculate the conditional and unconditional likelihood and probability of the chosen alternative on the test data and compare these to MNL, and to the case where only time and cost are distributed. Does the improvement in predictions justify the complexity of the model?

3. Try both normal and log-normal distributions for time and cost:
    - Which specification predicts better on the test data?
    - How many individuals have the wrong sign for time and cost in the sample if we use the normal distribution? (Check both the population distribution and the individual-specific parameters).