



## ΚΥΚΛΟΣ ΣΕΜΙΝΑΡΙΩΝ ΣΤΑΤΙΣΤΙΚΗΣ – ΟΚΤΩΒΡΙΟΣ 2016

### Gonzalo Garcia-Donato

*Department of Economy and Finance, Universidad de Castilla-La Mancha, Spain*

### Criteria for Bayesian model choice and implications on the $n \ll p$ problem

ΤΕΤΑΡΤΗ 26/10/2016  
15:00

ΑΙΘΟΥΣΑ 601, 6<sup>ος</sup> ΟΡΟΦΟΣ,  
ΚΤΙΡΙΟ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
(ΕΥΕΛΠΙΔΩΝ & ΛΕΥΚΑΔΟΣ)

#### ΠΕΡΙΛΗΨΗ

In model choice several statistical models are postulated as legitimate explanations for a response variable and this uncertainty is to be propagated in the inferential process. The type of questions one is aimed to answer is assorted ranging from e.g. identifying the 'true' model to produce more reliable estimates that takes into account this extra source of variability. Particular important problems of model choice are hypothesis testing, model averaging and variable selection. The Bayesian paradigm provides a conceptually simple and unified solution to the model selection problem: the posterior probabilities of the competing models. This is also named the posterior distribution over the model space and is a simple function of Bayes factors. Answering any question of interest just reduces to summarizing properly this posterior distribution.

Unfortunately, the posterior distribution may depend dramatically on the prior inputs and unlike estimation problems (where model is fixed) such sensitivity does not vanish with large sample sizes. Additionally, it is well known that standard solutions like improper or vague priors cannot be used in general as they result in arbitrary Bayes factors. Bayarri et al (2012) propose tackling these difficulties basing the assignment of prior distributions in objective contexts on a number of sensible statistical. This approach takes a step beyond a way of analyzing the problem that Jeffreys inaugurated fifty years ago.

In this talk the criteria is presented with emphasis on those aspects who serve to characterize features of the priors that, until today, have been popularly used without a formal justification. Originally the criteria were accompanied with an application to variable selection in regression models and here we see how they can be useful to tackle other important scenarios. In particular the high dimensional setting with  $n \ll p$  is analyzed under this perspective and the advantages of the proposed solution over the available methods is presented.



## AUEB STATISTICS SEMINAR SERIES – OCTOBER 2016

### Gonzalo Garcia-Donato

*Department of Economy and Finance, Universidad de Castilla-La Mancha, Spain*

### Criteria for Bayesian model choice and implications on the $n \ll p$ problem

WEDNESDAY 26/10/2016  
15:00

ROOM 601, 6<sup>th</sup> FLOOR,  
POSTGRADUATE STUDIES BUILDING  
(EVELPIDON & LEFKADOS)

#### ABSTRACT

In model choice several statistical models are postulated as legitimate explanations for a response variable and this uncertainty is to be propagated in the inferential process. The type of questions one is aimed to answer is assorted ranging from e.g. identifying the 'true' model to produce more reliable estimates that takes into account this extra source of variability. Particular important problems of model choice are hypothesis testing, model averaging and variable selection. The Bayesian paradigm provides a conceptually simple and unified solution to the model selection problem: the posterior probabilities of the competing models. This is also named the posterior distribution over the model space and is a simple function of Bayes factors. Answering any question of interest just reduces to summarizing properly this posterior distribution.

Unfortunately, the posterior distribution may depend dramatically on the prior inputs and unlike estimation problems (where model is fixed) such sensitivity does not vanish with large sample sizes. Additionally, it is well known that standard solutions like improper or vague priors cannot be used in general as they result in arbitrary Bayes factors. Bayarri et al (2012) propose tackling these difficulties basing the assignment of prior distributions in objective contexts on a number of sensible statistical. This approach takes a step beyond a way of analyzing the problem that Jeffreys inaugurated fifty years ago.

In this talk the criteria is presented with emphasis on those aspects who serve to characterize features of the priors that, until today, have been popularly used without a formal justification. Originally the criteria were accompanied with an application to variable selection in regression models and here we see how they can be useful to tackle other important scenarios. In particular the high dimensional setting with  $n \ll p$  is analyzed under this perspective and the advantages of the proposed solution over the available methods is presented.