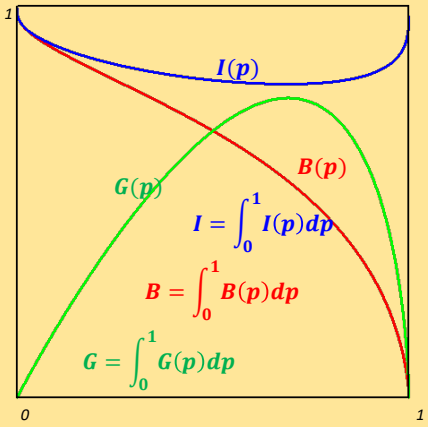
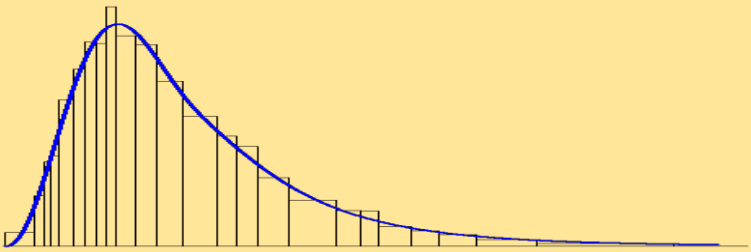
International conference on:

Distributions and Inequality Measures in Economics





Milan, February 20-21, 2020

Book of Abstracts

Department of Statistics and Quantitative Methods DISMEQ

Department of Economics, Management and Statistics DEMS

Statistics & Applications

SIS Società Italiana di Statistica

Istat Istituto Nazionale di Statistica

DIVERSITY IN SOCIO-ECONOMIC GROWTH AT COUNTRY LEVEL: A MULTIVARIATE HIDDEN MARKOV MODEL

Silvia Pandolfi * Francesco Bartolucci † Fulvia Pennoni** Alessio Serafini‡

We use data derived from the World Bank and UNESCO Institute for Statistics, which are referred to 217 countries followed for 18 years, from 2000 to 2017, to study the persistence of economic conditions at country level over time. We consider several variables, such as GDP per capita, educational levels, life expectancy at birth, infant mortality rate, school enrollment, government expenditure in education, and the value of the Gini index that measures the inequality level of the distribution of income through household survey data (Schultz, 1998). These variables are related to the human development index proposed by the United Nations Development Programme (1990) for measuring the well-being at country level. The proposed approach allow us to characterize disparities among countries in a dynamic fashion so as to evaluate the changes in inequality over time by estimating a multivariate hidden Markov (HM) model (Bartolucci, Farcomeni, and Pennoni, 2013; Zucchini, MacDonald, and Langrock, 2016), also known as latent Markov model. In particular we provide a dynamic clustering of the countries on the basis of the observed economic variables and, in turn, this permits to analyze inequalities in the development level. The resulting clustering is obtained by relying on a model tailored for longitudinal data rather than by predetermined threshold values as currently adopted by many international organizations to classify countries.
The HM model assumes the existence of an unobservable process, which follows a Markov chain with a discrete number of latent states, affecting the distribution of the observed variables. In particular, we consider multivariate continuous responses that, for the same time occasion, are assumed to be correlated according to a specific variance-covariance matrix, even conditionally on the latent states. In such a context, maximum likelihood estimation of model parameters is straightforward by means of the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977). Missing observations represent a relevant problem in this analysis, where values for some countries may be not available. Therefore, in this work, we also propose an approach for inference with missing data based on modifying the standard steps of the EM algorithm in a suitable way. In particular, we assume a structure of missing at random (MAR) responses (Little and Rubin, 2002) where the missing patterns are independent of

*Department of Economics, University of Perugia, Via A. Pascoli, 20 - 06123 - PERUGIA (e-mail: silvia.pandolfi@unipg.it)
†Department of Economics, University of Perugia, Via A. Pascoli, 20 - 06123 - PERUGIA (e-mail: francesco.bartolucci@unipg.it.)
**Department of Statistics and Quantitative Methods - University of Milano Bicocca - via Bicocca degli Arcimboldi, 8, 20126 MILANO (e-mail: fulvia.pennoni@unimib.it).
‡Department of Economics, University of Perugia, Via A. Pascoli, 20 - 06123 - PERUGIA (e-mail: alessio.serafini@unipg.it)

the missing responses given all the observed data. The resulting EM algorithm provides exact maximum likelihood estimates of the model parameters.
Since the resulting latent states may be ordered according to the average values of the economic variables, inequalities among countries are detected for each year on the basis of the estimated posterior probabilities of the countries of belonging to each latent state. In this way, we also explore patterns of disparities among countries and track their evolution from a dynamic perspective.

Keywords: *continuous outcomes, countries inequalities, dynamic clustering, missing data*

REFERENCES

Bartolucci F., Farcomeni A., and Pennoni F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences, Taylor & Francis.

Dempster A.P., Laird N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, **39**, 1-38.

Little R.J.A. and Rubin D.B. (2002). *Statistical Analysis with Missing Data*, 2nd Edn. Wiley Series in Probability and Statistics, Wiley

Schultz T.P. (1998). Inequality and the Distribution of Personal Income in the World: How It Is Changing and Why. *Journal of Population Economics*, **11**, 307-344.

Zucchini W., MacDonald I.L., and Langrock R. (2016). *Hidden Markov Models for Time Series: An Introduction using R*. CRC press.

United Nations Development Programme (1990). *Human Development Report 1990*. Oxford: Oxford University Press.