Università di Pisa
Sant'Anna Scuola Universitaria Superiore Pisa
Consiglio Nazionale delle Ricerche
Università degli Studi di Cagliari

Book of Short Papers

SIS 2021

SIS 2021
PISA, 21-25 GIUGNO 2021

SIS Società Italiana di Statistica

Editors: Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891927361

Contents

Cira Perna, Università degli Studi di Salerno (Chair of the Program Committee)
Nicola Salvati, Università di Pisa (Chair of the Local Organizing Committee)
Francesco Schirripa Spagnolo, Università di Pisa

Program Committee: Cira Perna (Chair), Massimiliano Caporin, Paola Cerchiello, Fabio Divino, Leonardo Egidi, Raffaele Guetto, Francesco Lagona, Silvia Loriga, Francesca Lotti, Mariagiulia Matteucci, Alessio Pollice, Maria Giovanna Ranalli, Nicola Salvati, Luca Secondi, Isabella Sulis, Luca Tardella, Donatella Vicari, Domenico Vistocco

Local Organizing Committee: Nicola Salvati (Chair), Gaia Bertarelli, Bruno Cheli, Paolo Frumento, Fosca Giannotti, Caterina Giusti, Piero Manfredi, Stefano Marchetti, Lucio Masserini, Vincenzo Mauro, Barbara Pacini, Dino Pedreschi, Francesco Schirripa Spagnolo, Chiara Seghieri.

Technical Committee: Luca Frigau (Chair), Silvia Columbu, Giulia Contu, Mara Manca, Marco Ortu, Maurizio Romano, Cristian Usala, Gianpaolo Zammarchi.

Organizers of Specialized and Solicited Sessions: Giada Adelfio, Emanuele Aliverti, Michela Battauz, Marco Bee, Mauro Bernardi, Luisa Bisaglia, Claudio Ceccarelli, Claudio Conversano, Marco Corazza, Giovanni Battista Dagnino, Antonella D'Agostino, Giovanni D'Alessio, Francesco Finazzi, Michela Gnaldi, Luigi Grossi, Roberto Impicciatore, Giovanna Jona Lasinio, Brunero Liseo, Orietta Luzi, Lucio Masserini, Letizia Mencarini, Stefania Mignani, Gianna Monti, Nazareno Panichella, Pierfrancesco Perri, Lea Petrella, Alessandra Petrucci, Paolo Righi, Elena Stanghellini, Venera Tomaselli, Cristina Tortora, Rosanna Verde, Paola Vicard, Maria Prosperina Vitale, Alberto Zazzaro.

Causes of death patterns and life expectancy: looking for warning signals

Cause di morte e speranza di vita: alla ricerca di segnali d'avvertimento

Stefano Mazzuco, Emanuele Aliverti, Daniele Durante and Stefano Campostrini

Abstract The evolution of longevity across countries is quite diverse and it still remains unclear what determined such different patterns throughout the last decades. In this paper we consider a Bayesian nonparametric mixture of B-splines for life expectancy trajectories that characterizes locally-varying similarities across functions, learning where country-specific trajectories are likely to overlap and where instead they tend to diverge. A preliminary comparison among Italy and United States indicates interesting trends in the evolution of life expectancy, with trajectories that overlap until the early 80s and then diverge substantially. We attempt to justify such differences by studying variations in the causes of premature mortality across these periods of interest, trying to justify potential driving factors for such divergences.
Abstract *Vi sono evidenti diversità nell'evoluzione della longevità tra paesi, ma non è ancora chiaro cosa abbia determinato andamenti così diversi negli ultimi anni. In questo lavoro, si considerano i dati sulle cause di morte per spiegare i diversi patterns di mortalità nel tempo, utilizzando un modello Bayesiano nonparametrico basato su misture di basi B-splines per caratterizzare similarità locali tra le diverse traiettorie di longevità nel tempo. Tale approccio consente di individuare in quali finestre temporali la longevità è simile tra paesi e in quali periodi diverge. Un primo confronto tra Italia e Stati Uniti offre dei primi risultati rilevanti che indicano l'inizio degli anni 80 come soglia da cui le differenze di longevità tra i due paesi sono divenute statisticamente rilevanti. Le motivazioni legate a tali differenze sono ricercate confrontando le traiettorie delle cause di morte.*

Key words: Bayesian nonparametrics, causes of death, life expectancy.

Stefano Mazzuco
Università degli studi di Padova, e-mail: stefano.mazzuco@unipd.it

Emanuele Aliverti
Università Ca' Foscari di Venezia, e-mail: emanuele.aliverti@unive.it

Daniele Durante
Università Bocconi di Milano, e-mail: daniele.durante@unibocconi.it

Stefano Campostrini
Università Ca' Foscari di Venezia, e-mail: stefano.campostrini@unive.it

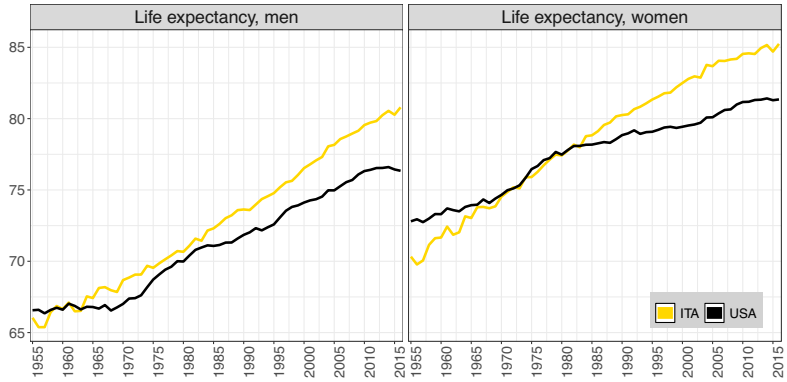

Fig. 1: Life expectancy in Italy and USA. Source: Human Mortality Database

1 Introduction

In the economically more advanced countries, longevity has steadily increased in the last decades. Comparing, for example, life expectancy at birth (e_0) in Italy and USA, we can observe in Figure 1 how these countries show comparable trends until the early 80s, resulting afterwards in substantially diverging trajectories. This result is even more striking if we consider that USA invests relatively almost twice in health of its resources (17% of GDP vs 8.7% of Italy; see [7]) and suggests the need of an improved understanding of the mechanism behind different patterns of longevity evolution, especially in relation to the mortality structure.

For example, Bergeron–Boucher et al [2] have recently shown that the extension of longevity is usually accompanied by a diversification of the causes of death. More specifically, Woolf and Schoomaker [13] analyze the trend of causes of death in USA, finding that midlife mortality caused by drug overdoses, alcohol abuses, suicides, and a list of organ system diseases have particularly increased in the last years. However, this finding has been contested (see Mehta et al, [6], who argue that cardiovascular diseases are the main responsible of US life expectancy stagnation). Such a controversy reflects the issue when dealing with cause-specific mortality, related with a competing risk setting: a cause-specific mortality rate can decline because there has been a significant improvement in treatment and/or prevention of that disease or just because other causes have grown meanwhile. Therefore, if we want to analyze the time trend of causes of death we need to take into account this feature. Stefanucci and Mazzuco [11] propose to combine Functional Data Analysis (FDA — see [10]) with Compositional Data Analysis (CDA — see [1]; [4]), limiting to a descriptive analysis of causes of death patterns.

Here, we propose a model-based analysis aimed at inferring patterns in causes of deaths that precede life expectancy stagnation. Motivated by Figure 1, we consider a flexible Bayesian nonparametric mixture of B-splines to learn local similarities

across life expectancies, assessing in which temporal blocks life-expectancy curves tend to overlap, and where instead they diverge. Subsequently, we analyze causes of death around the intervals of interest, in order to highlight what aspects of mortality have changed more considerably in those crucial years and what evidence they provide in terms of variation of life expectancy.

2 Data and methods

Data are collected from the Human Mortality Database [5], that ensures high quality data on mortality profiles of different European and non-European countries. Specifically, we focus here on Australia, Austria, Belgium, Bulgaria, Canada, Switzerland, Czech-Republic, Denmark, Spain, Finland, France, Great-Britain, Hungary, Ireland, Island, Italy, Japan, the Netherlands, Norway, Portugal, Slovakia, Sweden and United States of America. Moreover, causes of death data are taken from WHO mortality database.

We conduct analysis on these $n = 23$ countries, considering sex-specific and age-adjusted rates over a time period of $T = 62$ years ranging from 1955 to 2016. We consider 8 classes of causes of mortality, namely infections, neoplasms (all cancers with the exception of lung cancer), lung cancer, endocrines diseases, circulatory diseases, respiratory diseases, digestive diseases and external causes.

To flexibly model life expectancy patterns across countries $i = 1,\dots,n$ and years $t = 1,\dots T$, we treat the trajectory of e_0 as a function $y_i(t)$ and, following standard practice in FDA, we decompose it as

$$y_i(t) = f_i(t) + \varepsilon_i(t), \quad \varepsilon_i(t) \overset{\text{iid}}{\sim} N(0,\sigma^2), \tag{1}$$

for each country $i = 1,\dots,n$ and year $t = 1,\dots T$, where $\varepsilon_i(t)$s are independent Gaussian errors and $f_i(t)$ is an underlying smooth function. To include this smoothness, while avoiding strong assumptions on the functional form of $f_i(t)$, we model such a trajectory via B-splines [3], letting

$$f_i(t) = \sum_{k=1}^{K} \beta_{ik}\mathbf{B}_k(t), \tag{2}$$

where $[\mathbf{B}_1(t),\dots,\mathbf{B}_K(t)]$ denotes a set of fixed quadratic B-splines basis functions shared across countries, while β_{ik} denotes the country-specific coefficient referred to the k-th basis. Hence, overlapping and diverging patterns in the functions $f_i(\cdot)$, $i = 1,\dots,n$ across time are only regulated by ties among the associated coefficients β_{ik}. Motivated by our goal of learning such structures, we adapt the strategy in [8] to incorporate grouping effects for the coefficients β_{ik} via a model-based clustering induced by the following Dirichlet process prior on the coefficients

$$\beta_{1k},\dots,\beta_{nk} \mid F \overset{\text{iid}}{\sim} F, \quad F \sim \text{DP}(\alpha, F_0), \quad F_0 \sim N(0,\eta^2) \tag{3}$$

where DP(α, F_0) denotes a Dirichlet process with concentration parameter α and base measure F_0 which is assumed to be a Gaussian distribution with mean 0 and variance η^2. The discreteness of the DP is particularly appealing for our purposes, since it implies positive probabilities of ties among the coefficients β_{ik}, inducing local-clustering across curves as a byproduct. More specifically, denoting as $\beta^{\star}_{(h)k}$, $h = 1, \ldots, H_k$ the $H_k \leq n$ distinct values of the B-splines coefficients for the k-th basis, then, if $\beta_{ik} = \beta_{jk} = \beta^{\star}_{(h)k}$, countries i and j are expected to exhibit overlapping life-expectancy trajectories in the interval associated with the k-th spline basis. This local clustering can be formally characterized by $S_{(h)k} = \{i : \beta_{ik} = \beta^{\star}_{(h)k}\}$, and, thus, inference on the partitions $\rho_k = \{S_{(1)k}, \ldots, S_{(H_k)k}\}$ provides deeper insights on the trends underlying life-expectancy, flexibly learning which curves tend to overlap within specific blocks and characterizing uncertainty of this process.

Prior specification is completed by specifying a conjugate Inverse-Gamma distribution of the parameter $\sigma^2 \sim$ Inv-Gamma(a_0, b_0), while posterior inference proceeds via a collapsed back-fitted Gibbs sampler, exploiting the Polya-Urn scheme of the DP and leveraging the additive representation of the B-splines basis.

3 Preliminary results

We conduct posterior inference using 3000 Gibbs samples after a burn-in of 1000, setting $\alpha = 1$, $\eta^2 = 10$, $a_0 = b_0 = 0.01$. Effective sample size and autocorrelation plots did not provide evidence against convergence of the chains. In Figure 2 we obtain some preliminary results, focusing on the comparison between Italy and USA.

The first column of Figure 2 reports the smoothed estimated for the life expectancies and their associated credible intervals. Results indicate a common increasing trend for both states and sexes, with several important differences that are worth mention. In particular, for men we observe a period of overlapping curves in the late 50s, followed by separate trends in the late 60s and again a period of overlap during the late 70s and early 80s. Women instead report separate trends until the early 60s, followed by a long period of overlap until the late 80s and a subsequent increasing separation across the two curves.

These findings are further confirmed by the co-clustering probabilities, estimated as the proportion of MCMC sample in which Italy and United States are allocated in the same group. Such quantities are reported, as a function of time t, in the second column of Figure 2, while the third column of Figure 2 illustrates the ℓ_2 norm between the estimated life expectancy curves. Both panel indicate a common trend between men and woman after 1985, and quite different behavior in previous years.

Lastly, we compare the composition of the causes of death in the period of interest, focusing on premature mortality (≤ 70 years) in the window of interest $(1970 - 1990)$; see Figure 3. Such a composition has interestingly changed across this period of investigation, with the proportion of digestive system diseases and respiratory diseases showing markedly different trends across Italy and USA. Circulatory system diseases also report interesting trajectories, describing a notable

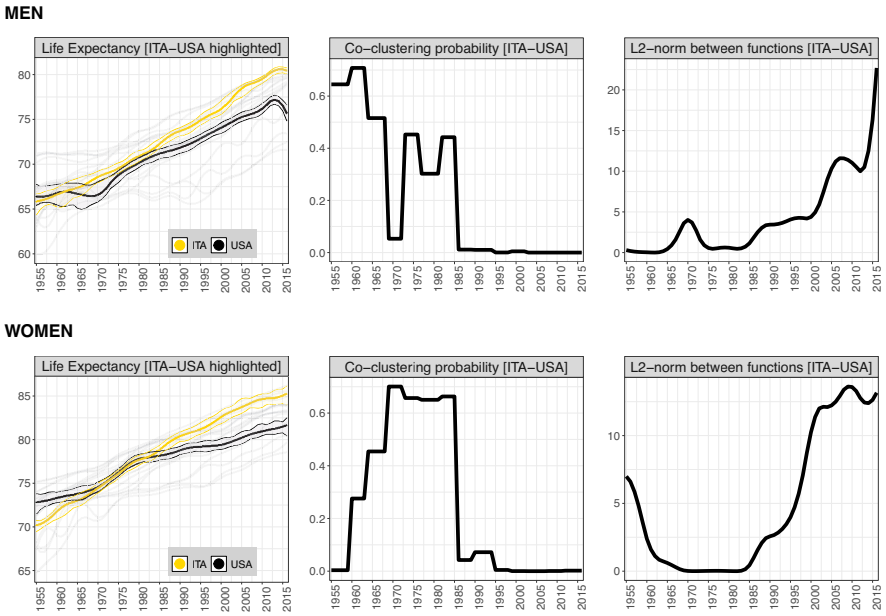

Fig. 2: First column: estimated functions $\hat{f}_i(t)$ via posterior mean and associated 95% credible intervals. Light gray curves represent all estimated countries, while Italy and United States are highlighted in yellow and black, respectively. Second column and third column focus on the comparison between Italy and United States, and depict co-clustering probability and ℓ_2 norm between estimated functions $\hat{f}(\cdot)$, respectively.

improvement for Italian women compared to Americans'. Worth to be mentioned is also the evolution of infectious diseases and the peak associated with AIDS, particularly severe in the American male population.

4 Discussion

In this article, we have provided a first step toward understanding the differences in longevity between Italy and the USA. We propose a flexible Bayesian nonparametric model to learn local-similarities across life expectancy trajectories, locating where these curves begin to diverge and exploring the composition of the causes of death around such period.

These preliminary evidences can be interpreted as *signals* of the evolving changes, more than explicit causes of the underlying processes. In fact, a proper analysis of these incredibly complex phenomena should also take into account other factors influencing the overall mortality level, along with its composition. Some potential determinants are the evolution of obesity trends and, more importantly, income in-

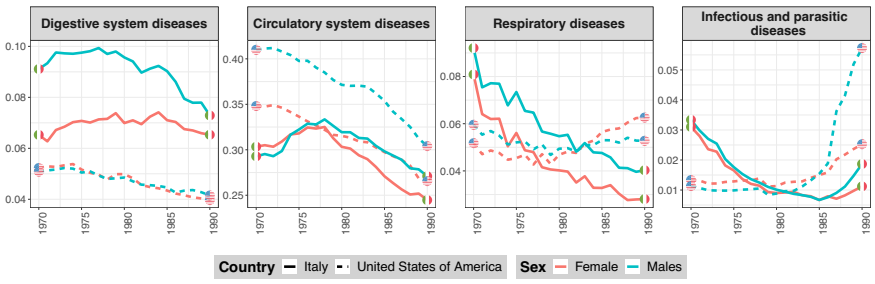

Fig. 3: Causes of death - premature mortality

equalities, which report substantial discrepancies in the period of interest [9]. Including the effects of this crucial determinants in our model will be the focus of future works, currently under investigation.

References

1. Aitchinson, J. (1986) *The Statistical Analysis of Compositional Data*. Chapman & Hall.
2. Bergeron–Boucher M-P, Aburto JM, van Raalte A. (2020) "Diversification in causes of death in low-mortality countries:emerging patterns and implications". *BMJ Global Health2020*;5:e002414. doi:10.1136/bmjgh-2020-002414
3. De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-verlag New York.
4. Egozcue, J. J. and Pawlowsky–Glahn, V. (2011) *Compositional Data Analysis: Theory and Applications*. John Wiley & Sons, Ltd.
5. Human Mortality Database (2020). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).
6. Mehta NK, Abrams LR, Myrskyla M. "US life expectancy stalls due to cardiovascular disease, not drug deaths". *Proceedings National Academy of Sciences* USA. 2020 Mar 31;**117(13)**:6998–7000. doi: 10.1073/pnas.1920391117.
7. OECD Health Statistics (2020), available at http://www.oecd.org/els/health-systems/health-data.htm
8. Paulon, G., Llanos, F., Chandrasekaran, B., and Sarkar, A. (2020). "Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults"., *Journal of the American Statistical Association*, 1–14
9. Raleigh, V.S. (2019). *Trends in life expectancy in EU and other* OECD *countries: Why are improvements slowing?*
10. Ramsay, J. O. and Silverman, B. W. (2005) *Functional data analysis*, vol. 2nd Ed. New York:Springer
11. Stefanucci M., Mazzuco S. (2020) "Analyzing Cause-Specific Mortality Trends using Compositional Functional Data Analysis". *Arxiv*:2007.15896.
12. Stringhini, S., Carmeli, C., Jokela, M., Avendaño, M., Muennig, P., Guida, F., Ricceri, F., d'Errico, A., Barros, H., Bochud, M. and Chadeau-Hyam, M. (2017) *Socioeconomic status and the* 25×25 *risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women* The Lancet, 389(10075), pp.1229-1237.
13. Woolf, S. H. and Schoomaker, H. (2019) "Life Expectancy and Mortality Rates in the United States, 1959–2017". *Journal of the American Medical Association*,**322**, 1996–2016