

Estimación de un modelo computacional mediante computación Bayesiana Aproximada

Juan Ignacio Baccino Costa , Mauro Loprete , Alvaro Valiño , Daniel Ciganda

Abstract El primer objetivo de este trabajo es estimar los parámetros de un modelo computacional del comportamiento reproductivo, en ausencia de una expresión analítica para la función de verosimilitud. Para esto, se utilizan técnicas de Computación Bayesiana Aproximada (ABC) y se analiza la incertidumbre asociada a las predicciones del modelo. Se trabaja con el modelo *Comfert*, un modelo de microsimulación que modela las trayectorias reproductivas de una cohorte de mujeres en un régimen de fecundidad natural, es decir, en ausencia de intentos dirigidos a prevenir nacimientos. Con estas trayectorias simuladas se obtienen las tasas específicas de fecundidad por edad para dicha cohorte y se utiliza esta información para ajustar el modelo a las tasas observadas en una población histórica. Los datos utilizados provienen de una cohorte de Hutteritas, una comunidad anabaptista frecuentemente estudiada en demografía por su rechazo del uso de métodos anticonceptivos y el nivel elevado de su fecundidad. Por otro lado, como objetivo secundario, se plantea el estudio sobre la incertidumbre en la predicción del modelo anteriormente descrito. Para ello, se realizó primero una estimación puntual considerada óptima (aquella que minimiza el Error Cuadrático Medio) y luego se construyeron diferentes intervalos de credibilidad. Si bien todos se encuentran contruidos al 95 % de credibilidad, la incertidumbre varía notablemente en función de los niveles de tolerancia considerados en el algoritmo. Los resultados obtenidos ilustran la utilidad del enfoque bayesiano en la realización de inferencia estadística sobre modelos computacionales, al igual que una relación positiva entre el nivel de incertidumbre y el incremento en la tolerancia del algoritmo.

Palabras clave: Modelos Computacionales - Computación Bayesiana Aproximada - Fecundidad - Error Cuadrático Medio - Intervalos de credibilidad

Introducción

El presente trabajo tiene como principal objetivo utilizar el enfoque ABC para estimar un modelo demográfico de microsimulación para el que no es posible derivar una expresión de la función de verosimilitud, pero desde el que es posible simular datos que pueden compararse con los datos provenientes de una población histórica. Por otro lado, como objetivo secundario, se plantea el estudio sobre la incertidumbre en la predicción del modelo anteriormente mencionado.

El modelo con el que se trabajó, *Comfert* (Ciganda and Todd 2019), simula las trayectorias reproductivas de una cohorte de mujeres en un régimen de fecundidad natural (población en la que no existe control explícito de la natalidad).

Para ajustar el modelo se utilizan una serie de tasas específicas de fecundidad por edad de una cohorte de mujeres de la población Hutteritas, una comunidad religiosa Anabaptista con un estilo de vida tradicional que incluye el rechazo al uso de métodos anticonceptivos.

Se entiende como proceso reproductivo a la secuencia de nacimientos de una mujer y edades de la madre al nacimiento.

A partir de esta secuencia es posible calcular los indicadores del nivel de la fecundidad más frecuentemente utilizados, como las tasas específicas de fecundidad por edad, definidas como:

$${}_nF_x [0, T] = \frac{{}_nB_x [0, T]}{{}_nL_x [0, T]}$$

siendo el numerador la cantidad de nacimientos de mujeres de edades entre x y $x + n$ sobre la población promedio (en T años) de este grupo de mujeres. Como se puede ver, esta tasa mide la cantidad promedio de hijos por mujer a la edad x .

Técnicas utilizadas

- Modelo de microsimulación
- Computación Bayesiana Aproximada ABC

El modelo de microsimulación, *Comfert* esta enteramente programando en R, al ser un modelo de simulación de tiempos discretos, utilizamos la librería *lubridate*, lo cual nos permite trabajar con los tiempos de los eventos principales (Nacimiento, Edad de Unión, Tiempo de no susceptibilidad, etc) de una trayectoria individual, dado el tipo de modelos se necesita una gran eficiencia velocidad en la ejecución de las simulaciones, es por esto, que se hace ayuda de el paquete *data.table* (Dowle and Srinivasan 2021).

En base de las características del modelo construido, sus parámetros son variables aleatorias con diferentes distribuciones que determinan el comportamiento de los individuos y es **imposible** encontrar una expresión analítica de la función de verosimilitud, es que se recurren a técnicas de Computación Bayesiana Aproximada (ABC).

Es por esto, que para implementar el algoritmo, fue necesario utilizar librerías como *mlgcp* (Dancik and Dorman 2008) para poder implementar procesos de simulación gaussiana para explorar el espacio paramétrico de nuestro meta-modelo.

Comenzando con distribuciones a priori de los parámetros a estimar el modelo y luego mediante sucesivas simulaciones del modelo ejecutadas en paralelo, mediante la librería *parallel* (R Core Team 2021) del modelo **Comfert** aproximamos la distribución a posteriori marginal de los mismos que mejor ajustan a los observados.

A continuación, a modo de resumen los principales parámetros del modelo *Comfert* y el rol de los mismos en el algoritmo **ABC**.

Parámetro	Descripción	Rol ABC	Valor
α	Edad de inflexión del declive de la fecundidad	A estimar	30-40
κ	Tasa de declive de la fecundabilidad	A estimar	0.20-0.4
shape	Heterogeneidad de la Fecundidad (entre mujeres)	Fijo	5
rate	Heterogeneidad de la Fecundidad (entre mujeres)	Fijo	38
μ	Edad media de la unión	A estimar	15-20
σ	Desvío de la edad media de la unión	A estimar	0.1-0.30
nsp	Período de no susceptibilidad	Fijo	6
srb	Ratio por sexo al nacimiento (en términos de hombres)	Fijo	0.515

Cuadro 1: Resumen

Por último, a la hora de estudiar la incertidumbre en la predicción del modelo, se procedió a computar el ECM en cada combinación de α y κ seleccionando como candidatos de valores pertenecientes a la estimación de la distribución a posteriori aquellos que tienen error menor o igual al percentil que acumula un 10 %, 50 %, y 75 % de la probabilidad de la distribución del error respectivamente. A continuación se muestra una visualización de los diferentes valores mediante visualizaciones en *ggplot2* (Wickham 2016)

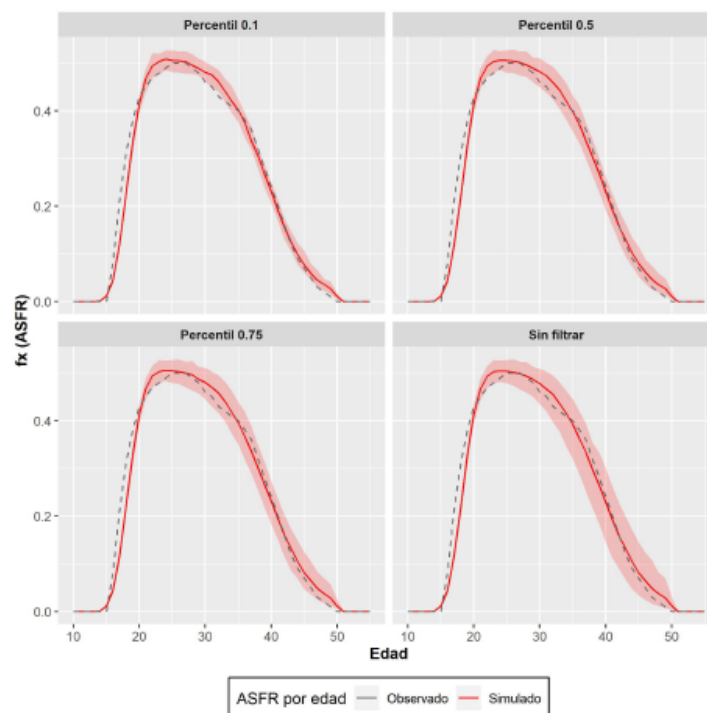


Figura 1: Gráfico de líneas de las ASFR, valores observados, mediana e intervalo de confianza al 95 %. Se observa que la amplitud de los intervalos de confianza aumenta conforme se incrementa el valor de epsilon seleccionado y la edad de la madre.

Referencias

- 10 Carnell, Rob. 2020. *Lhs: Latin Hypercube Samples*. <https://CRAN.R-project.org/package=lhs>.
- Ciganda, Daniel, and Nicolas Todd. 2019. "The limits to fertility recuperation." MPIDR Working Papers, no. WP-2019-024. <https://doi.org/10.4054/MPIDR-WP-2019-024>.
- Dancik, Garrett M, and Karin S Dorman. 2008. "mleqp: Statistical Analysis for Computer Models of Biological Systems Using r." *Bioinformatics* 24 (17): 1967.
- Dowle, Matt, and Arun Srinivasan. 2021. *Data.table: Extension of 'Data.frame'*. <https://CRAN.R-project.org/package=data.table>.
- Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Gutmann, Michael U., and Jukka Corander. 2015. "Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models." <http://arxiv.org/abs/1501.03291>.
- Henry, Louis. 1961. "Some Data on Natural Fertility." *Eugenics Quarterly* 8 (2): 81–91. <https://doi.org/10.1080/19485565.1961.9987465>.
- Kostaki, Anastasia, and Peristera Paraskevi. 2007. "Modeling Fertility in Modern Populations." *Demographic Research* 16: 141–94. <https://doi.org/10.4054/demres.2007.16.6>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schmertmann, Carl. 2003. "A System of Model Fertility Schedules with Graphically Intuitive Parameters." *Demographic Research* 9: 81–110. <https://doi.org/10.4054/demres.2003.9.5>.
- Sheps, Mindel C. 1965. "An Analysis of Reproductive Patterns in an American Isolate." *Population Studies* 19 (1): 65. <https://doi.org/10.2307/2173165>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.

Juan Ignacio Baccino Costa
Universidad de la República, FCEA, IESTA
nacho.baccino.3323@gmail.com

Mauro Loprete
Universidad de la República, FCEA, IESTA

Alvaro Valiño
Universidad de la República, FCEA, IESTA

Daniel Ciganda
Max Planck Institute for Demographic Research