UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Facultad de Estadística
Trabajo de Grado
Enero 2017

Estimación Bayesiana De Modelos Lineales Generalizados En Datos Funcionales.

Bayesian Estimation of Generalized lineals Models in Functional Data.

Leidy Tatiana Rugeles Diaz.[a]
leidyrugeles@usantotomas.edu.co

Wilmer Dario Pineda Ríos.[b]
wilmerpineda@usantotomas.edu.co

Resumen

Los estudios actuales, no siempre se deberían manejar como datos usuales donde un individuo representa un objeto en específico, si no como datos funcionales, tal que un individuo representa una curva, en la cual su naturaleza continua se conserva y se obtiene una reducción notable de la dimensión de los datos. Las variables funcionales se caracterizan por la evolución de una variable a lo largo del tiempo (proceso estocástico), de modo que los valores que toman son, en general, funciones de uno o varios argumentos en lugar de vectores como en análisis multivariante clásico. (*curso Máster en estadística, Universidad de Granada, 2016*). Dentro del análisis de datos funcionales se abordan los modelos lineales generalizados; estos modelos agrupan tanto modelos con variables respuestas numéricas como categóricas, lo que nos lleva a tener en cuenta otras distribuciones (Poisson, binomial, hipergeométrica, gamma, multinomial, etc.), además de la normal, y estos avances en la teoría de los modelos lineales se beben a Nelder y Wedderbum, y conjuntamente se impone el teorema de Bayes, el cual, trasciende la aplicación clásica, especialmente cuando se amplía a otro contexto en el que la probabilidad no se entiende exclusivamente como la frecuencia relativa de un suceso a largo plazo, sino como el grado de convicción personal acerca de que el suceso ocurra o pueda ocurrir (definición subjetiva de la probabilidad). Al admitir un manejo subjetivo de la probabilidad, el analista bayesiano podrá emitir juicios de probabilidad sobre una hipótesis H y expresar por esa vía su grado de convicción al respecto, tanto antes como después de haber observado los datos ([PDF]Estadística Bayesiana). Considerando estos temas se realiza un estudio de América Latina (Sin Surinam y Guyana) del índice de democracia EIU, a partir de tres índices, los cuales son: Producto Interno Bruto Per cápita y Índice de GINI desde el año 2002 hasta el 2015.

***Palabras clave*:** Índice EIU, Datos funcionales, base B-spline, modelos lineales generalizados, Estadística bayesiana.

Abstract

Current studies, should not always be handled as usual data where an individual represents a specific object, if not as functional data, such that an individual represents a curve, in which its continuous nature is conserved and a significant reduction of the dimension of the data. Functional variables are characterized by the evolution of a variable over time (stochastic process), so that the values they take are, in general, functions of one or several arguments instead of vectors as in classical multivariate analysis. (*Master course in statistics, University of Granada, 2016*). Within the analysis of functional data, generalized linear models are addressed; these models group both models with numerical and categorical variables, which leads us to take into account other distributions (Poisson, binomial, hypergeometric, gamma, multinomial, etc.), in addition to the normal one, and these advances in the theory of linear models are drunk to Nelder and Wedderbum, and jointly Bayes's theorem is imposed, which

[a]Estudiante de estadística Universidad Santo Tomás Bogotá
[b]Docente de estadística Universidad Santo Tomás Bogotá

transcends classical application, especially when it is extended to another context in which probability is not exclusively understood as the relative frequency of an event long term, but as the degree of personal conviction that the event occurs or may occur (subjective definition of probability). By admitting a subjective management of probability, the Bayesian analyst will be able to make probability judgments about a H hypothesis and express in this way his degree of conviction in this regard, both before and after having observed the data ([PDF] Bayesian Statistics). Considering these topics, a study of Latin America (Without Suriname and Guyana) of the democracy index EIU is made, based on three indices, which are: Per cápita Gross Domestic Product and GINI Index from the year 2002 until 2015.

***Keywords*:** EIU Index, functional data, B-spline basis, generalized linear models, Estadística bayesianaBayesian statistics.

1. Introducción

La presente tesis realiza un análisis mediante la aplicación bayesiana de modelos lineales generalizados en datos funcionales. Este trabajo corresponde al estudio del índice de democracia:$EIUINDEX$ para 11 países de América Latina desde el año 2002 al 2015; se tienen variables que se consideran que están asociadas al tema de los índices de democracia. El estudio se enfoca de manera funcional, observando las curvas (procesos estocásticos) por país, así describiendo el comportamiento de cada una de las variables. La doctrina de la seguridad democrática considera que si la democracia es deseable y si debe extenderse tanto como sea posible en todo el mundo, es principalmente porque garantiza la paz y la seguridad; muchos afirman que la mejor prueba de ello es que "las democracias no tienen guerras entre ellas mismas". Los índice utilizados en este estudio, son índices que exponen la importancia de poder explicar los índices de democracia en términos de otros índices según en un estudio de tesis realizado en el 2016. (Jonathan Ibáñez, Índice de democracia, 2016)

La importancia de abordar este tema en conjunto, es que se puede observar otra manera de aplicar la estadística, teniendo en cuenta el tipo de datos, usualmente en la aplicación de la estadística, no se sale de lo clásico, muy pocas veces se aplica la estadística bayesiana, con la que se podría llegar a obtener mejores resultados, o muchas veces nos complicamos con minería de datos cuando podríamos manejar los datos como datos funcionales, ahora bien, lo que se va a abordar en esta tesis nos permite identificar diferentes maneras de sacarle información a los datos de manera efectiva y eficaz.

En el transcurso de los años se han elaborado significativos estudios en el análisis funcional, estudios hechos por Silverman, Ramsay, Horváth, kokoszka, Möller, entre otros. Los datos funcionales han evolucionado de tal forma que cada uno de los temas de la estadística clásica se puede explicar y manejar en el análisis funcional, desde lo descriptivo hasta modelos o manejo de multivariada para curvas o procesos estocásticos. La teoría propia de lo funcional proviene de series temporales, y así mismo como tema de la estadística usual se puede implantar series de tiempo funcionales son incluidas dentro del entorno de procesos Hilbertvaluados. En este ámbito, existen resultados recientes sobre estimación para-métrica funcional, así como sobre extrapolación (ver Bosq[18], Guillas [42], Ruiz-Medina y Salmerón [82], Salmerón y Ruiz-Medina [84], entre otros). La estimación del modelo generalizado bayesiano aplicado a los procesos estocásticos, depende de la distribución de la familia exponencial de los datos a escoger y la determinación de la función de verosimilitud, para luego aplicar un algoritmo.

Lo que caracteriza este tipo de temas es la falta de experiencia en la vida practica y aun más es que se puede decir que es una recopilación de temas que tienen bastante aplicabilidad, pero, aun se hay escasez de teoría. En el caso de modelos definidos en tiempo y espacios continuos, así como a la formulación de funciones y núcleos de auto-correlación, en el caso de modelos definidos en tiempo discreto y espacio continuos,se analizarán diferentes modelos de interacción temporal desde la perspectiva funcional, es decir, considerando diferentes familias de operadores de auto-correlación, caracterizados en términos de sus propiedades espectrales puntuales (Román Salmerón Gómez, 2008).

Sin embargo, la inferencia bayesiana posterior no se ha convertido en una de las herramientas de inferencia estándar para los datos funcionales. Crainiceanu et al. (2009b) especulado que:

Las posibles razones de la falta actual de la metodología bayesiana en el análisis funcional podrían ser: (1)

la conexión entre los modelos funcionales y demás modelos conjuntos no se conocía; y (2) las herramientas de inferencia bayesiana se percibieron como innecesariamente complejo y difícil de implementar.

Con estudios previos se ha llegado a deducir que las excelentes propiedades de análisis bayesiano en el contexto del análisis de datos funcionales se deben a: (1) reducción de dimensionalidad, lo que conduce a las bases de proyección bajas dimensiones; (2) la representación modelo generalizado de modelos funcionales, que ofrece un enfoque modular para modelar la extensión; y (3) ortogonalidad de las principales bases de componentes, lo que podría contribuir a la excelente convergencia cadena y propiedades de mezcla.

Por lo tanto, lo que se propone hacer en este trabajo es aplicar métodos no frecuentes en la vida practica para llegar a conclusiones que podrían ser mejores de lo que usualmente se suelen llegar y decidir cuándo aplicarlos.

2. Justificación

Lo que se pretende con este tipo de estudios, es encontrar diversas herramientas que permiten un análisis optimo, uniendo tres temas como lo es bayesiana, modelos generalizados y datos funcionales Desarrollando así, estimaciones que sean muy eficientes y una visualización y análisis del tiempo, con un tema que está siendo de interés para el manejo más que todo de la minería de datos, como lo son los datos funcionales. Este estudio se realiza con el fin de observar el comportamiento de algunos índices económicos de América, en especial el índice de democracia EIU, el cual, es una medición hecha por la Unidad de Inteligencia de The Economist, a través de la cual se pretende determinar el rango de democracia en 167 países, de los cuales 166 son estados soberanos y 165 son estados miembros de las Naciones Unidas. De este índice que es construido para analizar todos los países del mundo, solo se tendrá en cuenta América Latina.

Las observaciones muestrales de una variable funcional son funciones que en la mayoría de los casos proceden de la observación temporal de una variable estadística (realizaciones de un proceso estocástico). Los datos funcionales aparecen en campos muy diversos de aplicación de la estadística como la economía, ciencias de la salud y medio ambiente, entre otros. En los últimos años han proliferado los trabajos de investigación en los que se generalizan las técnicas multivariantes al caso de datos funcionales, dando lugar a una parte de la estadística conocida como Análisis de Datos Funcionales (FDA).

En la práctica las funciones muestrales son observadas en un conjunto finito de puntos que pueden ser desigualmente espaciados y diferentes para los individuos muestrales. Por ello el primer paso en FDA es reconstruir la verdadera forma funcional de las curvas muestrales a partir de sus observaciones discretas. Uno de los métodos más usados en la práctica para aproximar las funciones muestrales consiste en representarlas en términos de bases de funciones y aproximar sus coeficientes mediante interpolación, en el caso de datos observados sin error, o mediante mínimos cuadrados, en el caso de datos ruidosos. Esta metodología proporciona buenas aproximaciones cuando las funciones básicas tienen esencialmente las mismas características que el proceso que genera los datos. En otro caso este método de aproximación no tiene control sobre el grado de suavización de la curva y puede llevar a aproximaciones poco precisas. Por lo que se pretende es dar a conocer el estudio de un método más potente de aproximar una función a partir de datos discretos(María del Carmen Aguilera Morillo, Granada, Septiembre de 2009).

Con el método bayesiano se pretende obtener estimaciones más optimas que los métodos clásicos. El teorema de Bayes da respuesta a cuestiones de tipo causal, predictivas y de diagnóstico. En las cuestiones causales queremos saber cuál es la probabilidad de acontecimientos que son la consecuencia de otros acontecimientos. En las cuestiones predictivas queremos saber cuál es la probabilidad de acontecimientos dada información de la ocurrencia de los acontecimientos predictores. En las cuestiones de tipo diagnóstico queremos saber cuál es la probabilidad del acontecimiento (o acontecimientos) causales o predictivos dado que tenemos información de las consecuencias. Para resumir, en las situaciones causales o predictivas desconocemos las consecuencias y tenemos evidencia de las causas. Por el contrario, en las situaciones de diagnóstico desconocemos las causas y tenemos evidencia de las consecuencias.

3. Objetivos

3.1. Objetivo General

Proponer un modelo lineal generalizado Bayesiano con respuestas correlacionadas y funciones predictoras, explicando el índice EIU en base del PIB per cápita y el índice de GINI.

3.2. Objetivos específicos

- Describir el comportamiento en el tiempo de los índices, de diferentes países de Sudamérica
- Construir un algoritmo que permita generar aleatorios de la función posterior.

4. Antecedentes

El análisis de datos funcional se ocupa de la modelización estadística de variables aleatorias que toman valores en un espacio de funciones (variables funcionales). Varias técnicas estadísticas estándares tales como la regresión, ANOVA o componentes principales, entre otros, han sido considerados desde el punto de vista funcional. En general, estas metodologías se centran en variables funcionales independientes e idénticamente distribuidas. Sin embargo, en varias disciplinas de las ciencias aplicadas, existe un gran interés en la modelización de datos funcionales espacialmente correlados. En particular, la mayoría de ellos están interesados en el modelado de datos funcionales espacialmente correlacionados. Este es el tema aquí tratado. En concreto, este proyecto trata la predicción de curvas, cuando se dispone de una muestra de las curvas de una región con continuidad espacial (*Análisis Geostadístico de datos funcionales, María José Ginzo Villamayor*).

Analiza datos de Imágenes de Resonancia Magnética funcional (fMRI). y se quiso saber qué áreas del cerebro están activas. se sabe que las áreas activas emiten señales mayores en respuesta a un estímulo que las que no lo están. Metodología: inferencia bayesiana. (*Métodos bayesianos para el análisis de imágenes de resonancia magnética funcional, Alicia Quirós Carretero, Raquel Montes Diez, Dani Gamerman, Universidad Rey Juan Carlos, 2008*)

Estimación en modelos lineales generalizados para datos funcionales a través de probabilidad penalizada. Se analiza en una regresión el establecimiento del vínculo entre una respuesta escalar y un predictor funcional mediante un Modelo Lineal Generalizado Funcional. Primero se da un marco teórico y luego se discute la identificabilidad del modelo. El coeficiente funcional del modelo se estima a través de la probabilidad penalizada con una aproximación spline. La tasa de convergencia L2 de este estimador se da bajo la suposición de suavidad sobre el coeficiente funcional. Los argumentos heurísticos muestran cómo estas tasas pueden mejorarse para algunos marcos particulares. (*Estimación en modelos lineales generalizados para datos funcionales a través de probabilidad penalizada, Hervé Cardot, Pacal Sarda, Academic Press, Inc. Orlando, FL, USA, 2008*)

5. Marco Teórico

5.1. Índices Económicos

Los índices económicos sirven para identificar la situación de una nación o región en un momento determinado. Por ejemplo, los indicadores más recientes confirman que la economía mundial sufre graves pérdidas debido a la

entrada en recesión de las principales economías del planeta. Al mismo tiempo, el costo de materias primas como el petróleo por el aumento de su demanda, han generado un deterioro en la economía mundial (*Indices Económicos Mundiales Y Situación Económica Mundial, Roberto Osorio, Instituto Nacional de San Rafael, 2013*).

5.1.1. Índice de Democracia de la Unidad de Inteligencia Económica (EIU)

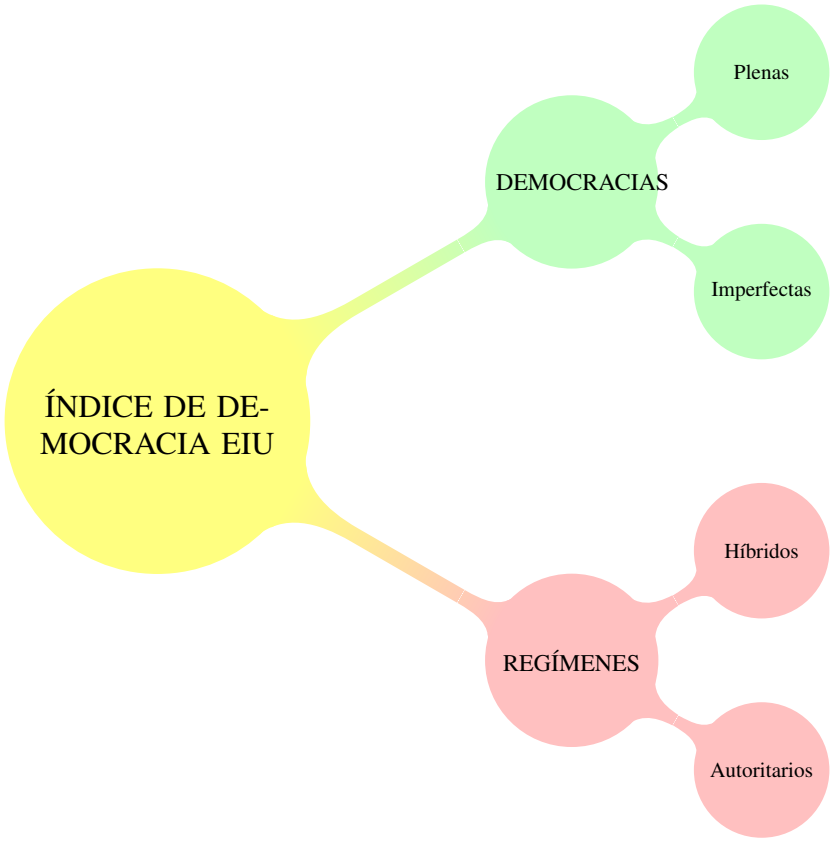
El Índice de Democracia (ID), calculado por The Economist Intelligence Unit (EIU), proporciona un reflejo del estado de la democracia alrededor del mundo. Sobre la base de puntuaciones de una serie de indicadores, cada país se clasifica en uno de cuatro tipos de régimen: "democracias plenas", "democracias imperfectas", regímenes híbridosz regímenes autoritarios"; definidos de la siguiente forma (EIU, 2015: p. 38):

- Democracias plenas [1]: países en los que no sólo se respetan las libertades políticas básicas y las libertades civiles, sino que también tienden a estar cimentados sobre una cultura política propicia para el florecimiento de la democracia. El funcionamiento del gobierno es satisfactorio. Los medios de comunicación son independientes y diversos. Existe un sistema eficaz de controles y equilibrios. El poder judicial es independiente y las decisiones judiciales se hacen cumplir. Solamente hay limitados problemas en el funcionamiento de las democracias
- Democracias imperfectas [2]: Estos países también tienen elecciones libres y justas, e incluso si hay problemas (como las infracciones a la libertad de los medios de comunicación), se respetan las libertades civiles básicas. Sin embargo, existen debilidades significativas en otros aspectos de la democracia, incluyendo problemas en la gobernanza, una cultura política poco desarrollado y bajos niveles de participación política.
- Regímenes híbridos [3]: Las elecciones tienen irregularidades sustanciales que a menudo previenen que sean libres y justas. La presión del gobierno sobre los partidos de la oposición y los candidatos puede ser común. Las debilidades graves en la cultura política, el funcionamiento del gobierno y la participación política son más frecuentes que en las democracias imperfectas. La corrupción tiende a ser generalizada y el estado de derecho es débil. La sociedad civil es débil. Normalmente existe hostigamiento y presión sobre los periodistas, y el poder judicial no es independiente.
- Regímenes autoritarios [4]: En estos países el pluralismo político está ausente o muy circunscrito. Muchos países de esta categoría son dictaduras absolutas. Pueden existir algunas instituciones formales de la democracia, pero éstas tienen poca sustancia. Las elecciones, en caso de que ocurran, no son libres ni justas. Hay indiferencia con respecto a los abusos y las violaciones de las libertades civiles. Los medios de comunicación usualmente son propiedad del Estado o están controlados por grupos conectados con el régimen gobernante. Existe represión de las críticas hacia el gobierno y hay censura generalizada. No existe un poder judicial independiente.

Con respecto a la definición de democracia, el último reporte destaca que a pesar de que los términos de libertad y democracia a menudo son utilizados de forma intercambiable, los dos no son sinónimos. La democracia puede ser vista como un conjunto de prácticas y principios que institucionalizan y por ende, en última instancia, protegen la libertad (EIU, 2015: p.34).

La principal fuente de información que utiliza la Economist Intelligence Unit (EIU) es la evaluación de expertos, aunque no se suministra ninguna información detallada sobre los expertos consultados ni la temporalidad de las consultas. Se indica que se utilizan encuestas de opinión para los países que cuentan con ellas. La principal encuesta de opinión que consideran es la World Values Survey (WVS). 4 Indicadores basados en encuestas predominan en las categorías de participación política y cultura política, y son utilizadas poco en las categorías libertades civiles y funcionamiento del gobierno. Adicionalmente a la World Values Survey, otras fuentes son utilizadas, como el Eurobarómetro, el Latinobarómetro, las encuestas de Gallup y encuestas nacionales. A partir de la calificación obtenida en el ID, se distribuyen los países en los cuatro tipos definidos de régimen, de la siguiente forma:

[1] Democracias plenas: calificaciones de 8 a 10.
[2] Democracias imperfectas: calificaciones de 6 a menos de 8.
[3] Regímenes híbridos: calificaciones de 4 a menos de 6.
[4] Regímenes autoritarios: calificaciones inferiores a 4.



5.1.2. Producto Interno Bruto (PIB) Per Cápita

El PIB per cápita, ingreso per cápita o renta per cápita es un indicador económico que mide la relación existente entre el nivel de renta de un país y su población. Para ello, se divide el Producto Interior Bruto (PIB) de dicho territorio entre el número de habitantes.

El empleo de la renta per cápita como indicador de riqueza o estabilidad económica de un territorio tiene sentido porque a través de su cálculo se interrelacionan la renta nacional (mediante el PIB en un periodo concreto) y los habitantes de este lugar.

El objetivo del PIB per cápita es obtener un dato que muestre de algún modo el nivel de riqueza o bienestar de ese territorio en un momento determinado. Con frecuencia se emplea como medida de comparación entre diferentes países, para demostrar las diferencias en cuanto a condiciones económicas.

"Este índice se calcula a partir del PIB y la población del país."

Es importante señalar que el PIB que suele emplearse a la hora de calcular ingresos per cápita es el expresado en términos nominales. En otras palabras, se utilizan precios vigentes de los bienes y los servicios producidos en dicho periodo y no constantes como son los del PIB real.

No obstante, es a menudo un ratio discutido debido a que no aporta la información suficiente al ignorar importantes aspectos como la desigualdad en el reparto de riquezas en los países, el factor de la educación o el nivel de desarrollo de dichos lugares. Aunque normalmente existe una relación directa entre el nivel de renta de un lugar y el nivel de aspectos como la sanidad, la educación y el desarrollo, no siempre la renta per cápita es capaz de mostrar de manera absoluta y veraz el auténtico nivel de vida de un ciudadano en un país determinado.

En ese sentido, a menudo se dice que esta magnitud no expresa bien la realidad en situaciones de desigualdad

o descontento social, especialmente en situaciones en las que la economía de un país crece pero esta mejora macroeconómica no siempre se refleja en la calidad de vida del ciudadano ni en su poder adquisitivo.

5.1.3. Índice de GINI

Éste coeficiente es una medida de concentración del ingreso, entre los individuos de una región, en un determinado periodo. Esta medida está ligada a la Curva de Lorenz . Toma valores entre 0 y 1, donde 0 indica que todos los individuos tienen el mismo ingreso y 1 indica que sólo un individuo tiene todo el ingreso.

Mide el grado de desigualdad de la distribución del ingreso o la desigualdad de la riqueza de una región. Y no mide el bienestar de una sociedad. Tampoco permite, por sí sólo, determinar la forma como está concentrado el ingreso; ni indica la diferencia en mejores condiciones de vida en un país u otro. El índice de Gini es el coeficiente de Gini expresado en referencia a 100 como máximo, en vez de como 1, y es igual al coeficiente de Gini multiplicado por 100. Una variación de dos centésimas del coeficiente de Gini (o dos unidades del índice) equivale a una distribución de un 7 % de riqueza del sector más pobre de la población (por debajo dela mediana) al más rico (por encima de la mediana).

Aunque el coeficiente de Gini se utiliza sobre todo para medir la desigualdad en los ingresos, también puede utilizarse para medir la desigualdad en la riqueza. Este uso requiere que nadie disponga de una riqueza neta negativa.

5.2. Análisis estadístico

5.2.1. Datos funcionales

Los datos funcionales son variables que evolucionan a lo largo del tiempo a lo cual se le conoce como proceso estocástico, de tal forma que los valores en este caso son funciones, mas no vectores como en el caso multivariado. Es decir, que al estudiar estos datos, los análisis se aplican a curvas o cualquier elemento que cambie sobre un espacio continuo.

$$\text{Datos funcionales = Las funciones de } X_i(t)$$

donde t es el tiempo continuo.

Al aplicar las medidas de tendencia central, dispersión y de relación a muestras de datos funcionales, se debe tener en cuenta que se está trabajando en un espacio vectorial L^2 (Funciones cuadradas integrables).

$$X : \Omega \mapsto L^2(\mathcal{T}) \text{ donde } t \in \mathcal{T} \text{ y asi mismo } X \in L^2(\mathcal{T})$$

Objetivos del análisis de datos funcionales

Expresar observaciones discretas derivadas de series de tiempo en forma de una función, representando toda la función de medida como una sola observación, para después de aplicar los conceptos estadísticos de análisis de datos multivariados, obtener estimaciones mas precisas de parámetros para uso en el análisis, además obteniendo una reducción efectiva del ruido de los datos a través de la suavización de curvas.

(Müller) Este método de análisis de datos funcionales puede extraer información adicional contenida en la función y sus derivados que normalmente no son disponibles con los métodos tradicionales de la estadística. Además, debido a que este enfoque trata esencialmente toda la curva como una sola observación, no hay preocupación acerca de las correlaciones entre las mediciones repetidas. Esto representa un cambio en la filosofía hacia el manejo de series de tiempo y datos correlacionados.

Características de los datos funcionales

1. Por lo general, no se hace suposiciones para-métricas sobre estos procesos.
2. Tienen suficientes datos para estimar un proceso suave subyacente a los datos.
3. No puede ser modelado por las fórmulas simples.
4. Puede no estar medido todo en los mismos puntos del tiempo.

Propiedades de los datos funcionales

La filosofía básica de análisis de datos funcionales es pensar en las funciones de datos observados como sujetos individuales, en lugar de simplemente como una secuencia de observaciones individuales. En la práctica, los datos funcionales observados generalmente se registran por separado y como n pares (t_j, y_j), donde y_j es una captura de la función en el tiempo t_j, posiblemente borrosa por error de medición. El tiempo es continuo, durante el cual los datos se registran funcionales, pero sin duda otros factores a parte de la continuidad pueden estar involucrados, tales como la posición espacial, la frecuencia, el peso, y así sucesivamente.

- **¿Que hace que los datos funcionales sean discretos?** Por lo general se quiere declarar que la función subyacente x son suaves, de modo que un par de valores de datos adyacentes, y_j y y_j+1 son necesariamente unidos entre sí en cierta medida y es poco probable que sean muy diferentes unos de otros. Si esta propiedad de suavidad no se aplicara, no habría nada que se pudiera obtener mediante el tratamiento de los datos como funcionales, más bien serian tan solo datos multivariantes.

  Suavizados, significa que la función x posee uno o más derivados, los cuales se indicaron por Dx, D^2x, y así sucesivamente, de modo que $D^m x$ se refiere a la derivada de orden m, y $D^m X(t)$ es el valor de dicho derivado en el argumento t. Usualmente se tendrá que usar el y_j datos discretos, donde $j = 1, ..., n$ para estimar la función de x y al mismo tiempo un cierto número de sus derivados. El modelamiento de las tasas de desplazamiento de un sistema, a menudo se llama el análisis de la dinámica de un sistema.

- **Muestras de datos funcionales** El registro o la observación de la función x_I podrían consistir en n_i parejas de (y_{ij}, t_{ij}), donde $j = 1, ..., n_i$. Puede ser que los valores de t_{ij} sean los mismos para cada registro, pero también puede variar de un registro a otro. Puede ser que el intervalo $\mathcal{T}$ sobre el cual se recogen los datos que también varían de un registro a otro.

  Por lo general, la construcción de las observaciones funcionales x_i, utilizando los datos discretos y_{ij} se lleva a cabo por separado o de forma independiente para cada registro i. No obstante, cuando el ruido es bajo, o los datos están escasamente muestreados o en número son pocos, puede ser esencial utilizar la información de vecinos o curvas similares para obtener estimaciones más estables de una curva específica.

  A veces, el tiempo t se considera de forma cíclica, por ejemplo, cuando t es el tiempo de años, esto significa que las funciones satisfacen condiciones del contorno periódicas, donde la función de x en el comienzo del intervalo de $\mathcal{T}$ recoge sin problemas los valores de x en el extremo. Así mismo cuando los datos para las funciones no satisfacen las condiciones del contorno, son llamados no periódica.

- **La interacción entre suave y la variación con ruido** La suavidad se da en el sentido de poseer un cierto número de derivados, es una propiedad de la función latente x, y puede no ser del todo evidente en el vector de datos inicial $y = (y_1, ..., y_n)$ debido a la presencia de error en las observaciones. Esto se expresa como:

  $$y_j = x(t_j) + \varepsilon_j$$

  donde el ruido del término j contribuye a una rugosidad de los datos iniciales. Una de las tareas en que representan los datos iniciales como funciones puede consistir en intentar filtrar este ruido tan eficientemente como sea posible. Sin embargo, en otros casos podemos hacer cumplir la estrategia alternativa de dejar el ruido en la función estimada.

La notación vectorial conduce a expresiones más simples, por lo que expresa el modelo de "señal más ruido"(modelo anterior) como:

$$y = x(t) + e$$

en la que y, $x(t)$, t y e son todos los vectores de la columna de longitud n.

La matriz de varianza-covarianza para el vector de los valores observados y es igual a la matriz de varianza-covarianza para el vector correspondiente a los valores residuales, ya que los valores $x(t_j)$ están considerados como efectos fijos con varianza 0. Sea Σ_e la notación para la matriz residual de varianza-covarianza, que expresa cómo los residuos varían en muestras repetidas que es idéntica en todos los aspectos salvo por el ruido o la variación de error.

- **Modelo estándar para el error** El modelo estadístico estándar para la distribución de la j es asumir que se distribuyen de forma independiente con media cero y varianza constante σ^2. En consecuencia, el modelo estándar es,

  $$Var(y) = \Sigma_e = \sigma^2 I$$

  donde la matriz identidad I es de orden n. Es posible que tenga que tomar en cuenta una correlación entre los vecinos de e_j. La auto-correlación que a menudo vemos en los residuos funcionales refleja el hecho de que la variación funcional que elegimos se ignora en sí misma, probablemente sin problemas a una escala más fina de la resolución.

  De hecho, el concepto de error distribuido de forma independiente en la norma modelo, que, a medida que n aumenta, se convierte en lo que se denomina ruido blanco, no es realista o realizables en la naturaleza, porque el ruido blanco requeriría energía infinita de lograr. Esto no significa necesariamente que será siempre esencial para modelar la varianza o estructura de auto-correlación en los residuos o errores. Tales modelos para Σ_e pueden quemar valiosos grados de libertad, reducir la velocidad de cálculo de manera significativa, y, finalmente, dar lugar a estimaciones de funciones que son prácticamente indistinguibles de lo que se logra por los supuestos de independencia de los residuos.

- **El poder de resolución de los datos**

  La frecuencia de muestreo o la resolución de los datos en iniciales (en bruto) es un factor determinante de lo que es posible en el modo de análisis de datos funcionales. Esto es esencialmente una propiedad local de los datos, y se puede describir como la densidad de los valores de los argumentos t_j en relación con la cantidad de curvatura en los datos, en lugar de simplemente el número de valores n de los argumentos. La curvatura de una función de x en el argumento t se mide generalmente por el tamaño de la segunda derivada, como se refleja en $|D^2x(t)$ o $[D^2x(t)]^2$. Donde la curvatura es alta, y es esencial disponer de suficientes puntos para estimar la función con eficacia. La suficiencia de esto depende de la cantidad de error ε_j; cuando el nivel de error es pequeño y la curvatura es leve, se puede llegar lejos con una baja tasa de muestreo.

  Cuando la función observada parece razonablemente suave, la velocidad de muestreo es alta, y el nivel de error es baja, se podría estar tentado a usar la primera diferencia hacia adelante $(y_{j+1} - y_j)/(t_{j+1} - t_j)$, o la diferencia central $(y_{j+1} - y_j - 1)/[(t_{j+1} - t_j - 1)]$, para estimar $Dx(t_j)$

  $$D^2x(t_j) \approx (y_{j+1} + y_{y-1} - 2y_j)/(\Delta t)^2$$

  Press et al. (1999) comentan sobre la forma sencilla de diferenciación para estimar los derivados que pueden salir mal, incluso cuando las funciones están disponibles analíticamente.

Representación de funciones por funciones de base

Un sistema de función de base es un conjunto de funciones conocidas como ϕ_k, los cuales son matemáticamente independientes entre sí y tienen la propiedad de que poder aproximar arbitrariamente bien cualquier función mediante la adopción de una suma ponderada o combinación lineal de un número suficientemente grande K de

estas funciones. El sistema de función de base más familiar es el conjunto de monomios que se utilizan para construir series de potencias,

$$1, t^2, t^3, ..., t^k$$

Los procedimientos de función de base representan una función de x mediante una expansión lineal

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

En efecto, los métodos de expansión base representan el mundo dimensional potencialmente infinito de funciones dentro del marco de dimensión finita de unos vectores como c. Por consiguiente, la dimensión de la expansión es K.

Una representación exacta o interpolación se consigue cuando $K = n$, en el sentido de poder elegir los coeficientes c_k para producir $x(t_j) = y_j$ para cada j. Por tanto, el grado en que los datos y_j se suavizan en oposición a la interpolación que se determina por el número K de funciones de base. Por lo que se ve a K en sí un como un parámetro que elegimos en función de las características de los datos.

Las funciones de base deben tener características que coinciden con los conocidos por pertenecer a las funciones que se estiman. Esto hace que sea más fácil de lograr una aproximación satisfactoria utilizando un número relativamente pequeño K de funciones de base.

El K es menor y cuanto mejor las funciones de base se reflejan ciertas características de los datos,

- Cuanto más grados de libertad, se tiene que probar las hipótesis y calcular intervalos de confianza exactos.
- Lo más probable es que los propios coeficientes pueden llegar a ser interesantes descriptores de los datos desde un punto de vista característico.

Estimación de la derivada:

$$D\widehat{x}(t) = \sum_{k}^{K} \widehat{x}_k \phi_k(t) = \widehat{c}' D\phi(t)$$

Sistemas de bases

- **Sistema de base de Fourier de los datos periódica**

  Puede que la expansión base más conocida es proporcionada por la serie Fourier:

  $$\widehat{x}(t) = c_0 + c_1 sin\omega t + c_2 cos\omega t + c_3 sin2\omega t + c_4 cos2\omega t + ...$$

  definida por la base $\phi_0(t) = 1$, $\phi_{2r-1}(t) = sinrwt$, y $\phi_{2r}(t) = cosrwt$. Esta base es periódica, y el parámetro ω determina el período $2\pi/\omega$. Si los valores de t_j están igualmente espaciados en $\mathcal{T}$ y el período es igual a la longitud del intervalo $\mathcal{T}$, entonces la base es ortogonal en el sentido de que la matriz de producto vectorial $\Phi'\Phi$ es diagonal, y se puede hacer igual a la identidad dividiendo las funciones de base por las constantes adecuadas, $\sqrt{n}$ para $j = 0$ y $n/2$ para el resto de j.

  La transformada rápida de Fourier (FFT) hace que sea posible encontrar todos los coeficientes extremadamente eficiente cuando n es una potencia de 2 y los argumentos son igualmente espaciados; en este caso podemos encontrar tanto los coeficientes c_k y todos los valores suaves n en $x(t_j)$.

- **Sistema de base de datos para spline indefinido**

Las funciones spline son la opción más común de sistema de aproximación para datos funcionales no periódicas o parámetros. Spline combinan el cálculo rápido de los polinomios con una flexibilidad mayor sustancialmente. Además, los sistemas básicos se han desarrollado para funciones spline que requieren una cantidad de cálculo que es proporcional a n, esta es una propiedad importante, ya que muchas aplicaciones implican miles o millones de observaciones.

- **Bases exponenciales y de poder**

  Este sistema consiste en una serie de funciones exponenciales

  $$e^{\lambda_1 t}, e^{\lambda_2 t}, ..., e^{\lambda_k t}$$

  donde todos los parámetros λ_k son distintas, a menudo $\lambda_1 = 0$. Las ecuaciones lineales diferenciables con coeficientes constantes, tienen como solución la expansión en términos de bases exponenciales.

  Donde las bases de poder se dan de la siguiente manera:

  $$t^{\lambda_1}, t^{\lambda_2}, ..., t^{\lambda_k}$$

  Esto es importante cuando t es estrictamente positivo, por lo tanto, las potencias negativas son posibles.

- **Bases polinómicas**

  La base de monomio es $\phi_k(t) = (t - \omega)^k$, y $k = 0, ..., K$ también es clásico, donde ω es un parámetro de desplazamiento que por lo general se elige para estar en el centro del intervalo de aproximación. Se debe tener cuidado para evitar errores de redondeo, ya que los valores de monomios son cada vez más altamente correlacionados a medida que aumenta el grado.

  Al igual que el desarrollo en serie de Fourier, los polinomios no pueden exhibir características muy locales sin necesidad de utilizar un gran K. Por otra parte, los polinomios tienden a encajar bien en el centro de los datos, pero exhiben un comportamiento muy poco atractiva en las colas. Son por lo general una mala base para la extrapolación o la previsión.

- **La base poligonal**

  Suavizar los datos en bruto (iniciales) observados no siempre es necesario, y sobre todo si nuestro interés no está en el ajuste a los datos en sí, sino más bien en que algún parámetro funcional no este conectado directamente a los datos. En los modelos lineales funcionales, podemos interpolar los datos con una base simple, y mover el tema de suavizado a donde pertenece, para saber la estimación del parámetro funcional. De hecho, los ataques de datos poligonales o lineales a trozos son muy recomendables, y pueden incluso ofrecer una estimación aproximada de la primera derivada.

- **La base en función de escalón**

  Los problemas de minería de datos a menudo implican un gran número de variables combinadas con tamaños de muestra fenomenales. Debido a las limitaciones computacionales pueden convertido en crítico, y se buscan los métodos más simples para trabajarlos. Una fracción de valores de variables puede ser visto como una transformación funcional con una base que consta de dos funciones escalonadas. Esta es, en efecto, un orden de un sistema de B-spline con un solo nudo interior. Una referencia reciente sobre la minería de datos en general, que tiene considerable material de clasificación basado en árboles es $Hastie, Tibshirani$ y $Friedman(2001)$.

- **La base constante**

  La función de base individual $1(t)$ cuyo valor es uno en todas partes. Es muy útil con sorprendente frecuencia. En primer lugar, proporciona un punto de referencia o hipótesis nula cuando se estiman funciones de coeficiente de regresión para el modelo lineal funcional y en otros lugares. En segundo lugar, es explícita en sistemas como el de Fourier, e implícitamente en las bases B-spline Por último, podemos ver una observación escalar como un dato funcional cuyo valor es igual en todas partes, y en consecuencia su valor se convierte en el coeficiente de la base constante.

- **bases empíricas y de diseño**

  Las bases de diseño también se pueden construir empíricamente mediante el análisis de componentes principales funcional. Estas bases tienen la propiedad de la optimización de la cantidad de variación en los datos explicada por los sistemas básicos de tamaño K. Si se desea que la sea lo mas compacta posible con el único objetivo de ajuste de los datos, el análisis de componentes principales es por lo general el método que se debería considerar en primer lugar.

Pasos para el análisis de datos funcionales

1. **Representación de datos (suavizado e interpolación):** Si un dato funcional i viene como un grupo de valores de medición discreta, $y_{i1}, y_{i2}, ..., y_{in}$, lo primero que se debe hacer es convertir estos valores a una función x_i con valores $x_i(t)$. Si estos valores discretos no tienen error, entonces se aplica un proceso de interpolación, pero si tienen algún error entonces la conversión de los datos a las funciones implica suavizamiento.

2. **Registro de datos y alineación:**

   - Dejar el intervalo de $\mathcal{T}$ sobre el cual, las funciones deben estar registradas $[T_1, t_2]$.
   - Se supone que cada x_i esta disponible mas allá de cada extremo de $\mathcal{T}$.
   - Se quiere los valores de:
     $$x_{ij} = x_i(t + \delta_i)$$
     donde los parámetros de desplazamiento δ_i alinean las curvas.
   - Las curvas x_{ij} son registrados por desplazamientos δ_i

   El registro de desplazamiento fue propuesto por Silverman en 1995. Es el método para estimar el tiempo de desplazamiento δ_i que transforman un sujeto de tiempo sincronizado. Silverman postula un simple caso de distorsión de tiempo, donde
   $$x_{ij} = x_i(t_{ij} - \delta_i) + \xi_{ij}$$

   Se propone estimar δ_i, $i = 1, .., m$ a través del siguiente algoritmo:

   - **Paso 1:** Elija un valor a partir de $\delta^{(0)} = \delta_1^{(0)}, ..., \delta_M^{(0)}$. Aunque por lo general la elección es $\delta^{(0)} = 0$ con un vector de 0 M-dimensional de cero entradas.
   - **Paso 2:** Para $v = 1, ...n$ calcular la media muestral de las funciones desplazadas
     $$\widehat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i(t_{ij} - \delta_i^{(v-1)})$$
     .
     Se actualiza el $\delta_i^{(v-1)}$, $i = 1, ..., M$ por
     $$\delta_i^{(v)} = \delta_i^{(v-1)} - \alpha \frac{\frac{\partial}{\partial \delta_i} REGSSE(\delta)|\delta = \delta_i^{(v-1)}}{\frac{\partial^2}{\partial \delta_i^2} REGSSE(\delta)|\delta = \delta_i^{(v-1)}}$$
     donde
     $$REGSSE(\delta) = \sum_{i=1}^{M} \int_0^{T_0} [x_i(t_{ij} - \delta_i) - \widehat{\mu}(t)]^2$$
     y $\alpha > 0$.
   - **Paso 3:** Repetir el Paso 2 hasta convergencia.

Una vez que se estima el desplazamiento, los tiempos observados pueden ser transformados mediante **la función de distorsión** $h_i(t) = t - \delta_i$.

3. **Trazados de pares de derivados:** Según Ramsay y Silverman, las pistas útiles a los procesos que dan lugar a los datos funcionales a menudo se pueden encontrar en las relaciones entre los derivados. Por ejemplo, dos funciones que muestran relaciones derivadas simples se encuentran con frecuencia como una fuerte influencia en los datos funcionales:

   La función exponencial $f(t) = C_1 C_2 e^{\alpha t}$, satisface la ecuación diferencial

   $$Df = -\alpha(f - C_1)$$

   y la sinusoidal $f(t) = C1 + C2sin[\omega(t - \tau)]$ con una constante de base τ que satisface

   $$D2f = -\omega^2(f - C_1)$$

   El trazado de la primera o segunda derivada contra el valor de la función explora la posibilidad de demostrar una relación lineal correspondiente a alguna ecuación diferencial. Por supuesto, normalmente, no es difícil de detectar estos tipos de variación funcional mediante el trazado de los datos propios. Sin embargo, El trazado del derivado superior contra el inferior a menudo es más informativo, en parte porque es más fácil de detectar desviaciones de la linealidad que de otras formas funcionales, y por otra parte debido a que la diferenciación puede exponer efectos que no se ve fácilmente en las funciones originales.

"El uso de derivados es importante tanto en la extención de rangos de los métodos de exploración de gráficos simples, y en el desarrollo de una metodología más detallada. La utilización de la información derivada en el estudio de los componentes de variación es implantada por un método llamado análisis diferencial director el cual identifica importantes componentes de la varianza de la estimación de un operador diferencial lineal. Los operadores diferenciales lineales, se estima a partir de datos construidos a partir de consideraciones de modelamiento externo, también desempeñan un papel importante en el desarrollo de métodos de regularización más generales que los de uso común."

Exploración de la variabilidad en los datos funcionales

Algunas formas de observar la variabilidad en los datos funcionales se dan a continuación:

- **Estadísticas descriptivas funcionales** Cualquier análisis de datos comienza con lo básico: los medios de estimación y las desviaciones estándar, pero de forma funcional. Pero algo importante para datos univariados y multivariados resulta ser no siempre tan simple para datos funcionales, ya que cada observación es una curva; Aquí se muestra que el registro o alineación (1.4.2) de la curva característica puede ser aplicada con el fin de separar variación de la amplitud de la variación de fase antes de utilizar estas estadísticas.

- **Análisis de componentes principales funcionales** La mayoría de los conjuntos de datos muestran un pequeño número sustancial de variación, incluso después de restar la función media de cada observación. Una aproximación a la identificación y exploración de estos, es adaptar el procedimiento multivariante clásico de análisis de componentes principales para datos funcionales. Algunas técnicas de suavizamiento o regularización se incorporan en el mismo análisis de componentes principales funcional, lo que demuestra que los métodos de suavizado tienen un papel mucho más amplio en el análisis de datos funcional que sólo en la etapa inicial de convertir observaciones discretas de forma funcional. Un análisis de componentes principales funcional puede hacerse más selectivo e informativo al considerar determinados tipos de variación de una manera especial.

Exploración y suavizamiento. de datos funcionales

- **Derivadas y integrales** La notación para la derivada de orden m de una función x es $D^m x$; esto produce fórmulas en un sentido más amplio $d^m x/dt^m$. Siendo la diferenciación un operador que actúa sobre una función de x para producir otra función Dx. Se refiere con $D^0 x$ a sí mismo. Se utiliza $D^{-1}x$ para referirse a la integral indefinida de x, ya que $D^1 D^{-1} x = D^0 x = x$.

  La integral definida se define como: $\int_a^b x(t)dt$.

- **Resumen de estadísticas para datos funcionales**

  1. **Moda y Media funcionales** Las estadísticas de resumen clásicas para datos univariados de la introducción estadística se aplican por igual a los datos funcionales.

  $$M_e = Lim_i + \frac{\frac{N}{2} - f_{i-1}(t)}{f_i(t)} * a_i(t)$$

  Esta es la mediana de las funciones. También se puede expresar la moda de la estadística clásica en términos de funciones, así:

  2. **Medias y varianzas funcionales** La media funcional de la muestra es un estimador del centro de la distribución funcional.

  $$\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t)$$

  Este es el promedio de las funciones a través de las repeticiones. Así mismo la función de la varianza es:

  $$Var_x(t) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t) - \bar{x}(t)]^2$$

  Y así mismo la desviación es la raíz de la varianza.

  3. **Función de covarianza y correlación** La covarianza funcional de la muestra es un estimador de la escala y la estructura de correlación de la distribución funcional.

  La función de covarianza resume la dependencia de los registros a través de diferentes valores de los argumentos, y se calcula para todos t_1 y t_2 por:

  $$Cov_x(t_1, t_2) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]$$

  Y la función de correlación es:

  $$Corr_x(t_1, t_2) \frac{Cov_x(t_1, t_2)}{\sqrt{Var_x(t_1) Var_x(t_2)}}$$

  4. **Función de covarianza cruzada y correlación cruzada**

  Si tenemos pares de funciones observadas (x_i, y_i), la forma en que estos dependen uno del otro puede ser cuantificada por la función de covarianza cruzada, así:

  $$Cov_{x,y}(t_1, t_2) = (N-1)^{-1} \sum_{i=1}^{N} [x_i(t_1) - \bar{x}(t_1)][y_i(t_2) - \bar{y}(t_2)]$$

  Y la función de correlación cruzada es:

  $$Corr_{x,y}(t_1, t_2) = \frac{Cov_{x,y}(t_1, t_2)}{\sqrt{Var_x(t_1) Var_y(t_2)}}$$

  Hay que tener en cuenta que $Corr_{x,y}(t_1, t_2) = Corry, x(t_2, t_1)$, estos son transpuestas uno de otro, en que cada uno es el reflejo de la otra sobre la diagonal principal $t_1 = t_2$. Aqui mismo se puede identificar si el proceso es estacionario o no.

Modelos lineales funcionales

Las técnicas clásicas de regresión lineal, análisis de varianza investigan la forma en la que la variabilidad en los datos observados se puede explicar por otras variables conocidas u observadas. Expresado como:

$$y = \mathbf{z}\beta + \varepsilon$$

donde, y por lo general es un vector de observaciones, β es un vector de parámetros, $\mathbf{z}$ es una matriz que define una transformación lineal del espacio de parámetros de observación, y ε es un vector de error con media cero. La matriz de diseño $\mathbf{z}$ incorpora covariables observadas o variables independientes.

5.2.2. Modelos lineales Generalizados

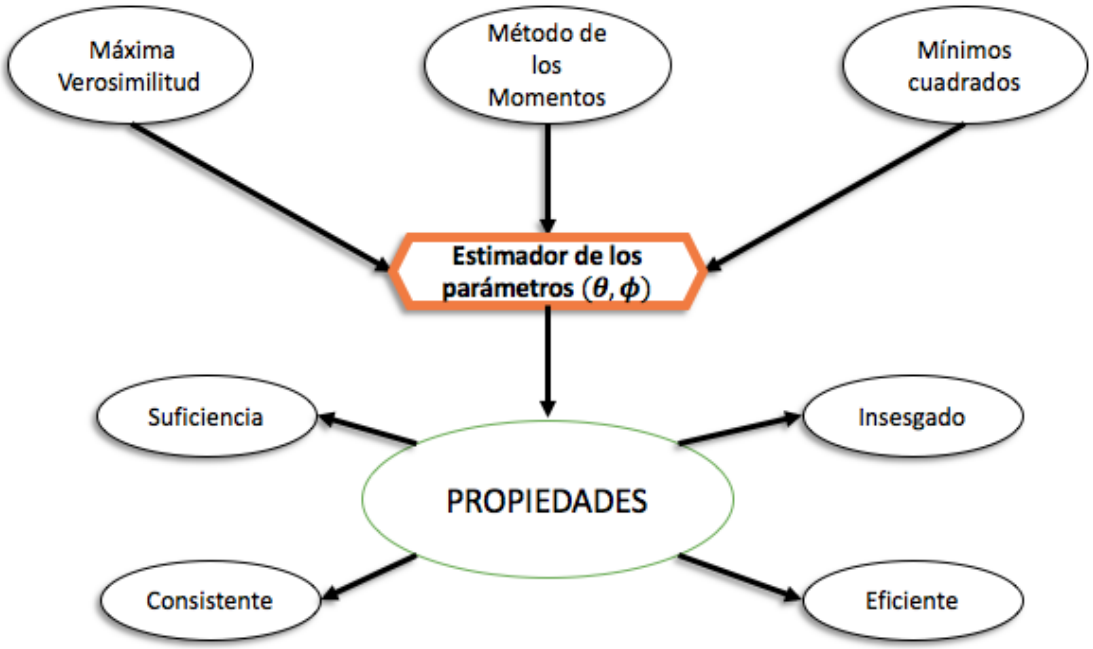
Los modelos lineales generalizados toleran que la variable dependiente tenga una distribución no normal, su utilización considera toda la familia exponencial, la cual depende del parámetro de dispersión y parámetro natural, y se expresa de la siguiente forma:

$$f(y, \theta, \phi) = exp\left\{\phi(y\theta - b(\theta) + c(y, \phi)\right\}$$

Donde, θ es el parámetro natural y ϕ es el parámetro de dispersión.

Es importante recalcar que la familia exponencial permite el cálculo de un estadístico suficiente, en este caso para θ y ϕ, ya que la distribución de probabilidad de la muestra representada por $f(y, \theta, \phi)$ permite la descomposición exponencial.

Realizando un paréntesis, para los análisis posteriores es necesario recordar los métodos de estimación puntual:



Componentes del modelo lineal generalizado

- **Componente aleatoria :** Corresponde a un vector aleatorio **Y**, la cual sigue una distribución de la familia exponencial, tal que $Y_1, Y_2, ..., Y_n$ son independientes e idénticamente distribuidos.Donde **Y** tiene distribución

$$exp\{\frac{y\theta - b(\theta)}{a(\theta)} + c(y, \phi)\}$$

Donde, θ es el parámetro natural, ϕ es el parámetro de dispersión y las funciones $c()$, $b()$ y $a()$.

$$\mu = E(Y) = b'(\theta)$$
$$Var(Y) = a(\phi)b''(\theta)$$

- **Componente sistemática :** Hace referencia al predictor lineal (η),

$$\eta_{nx1} = X_{nxk}\beta_{kx1}$$

- **Función de enlace ($g(\mu)$):** Relaciona la Componente Sistemática, con el valor esperados de la variable respuesta ($E(Y/X) = \mu$), a través de la función:

$$g(\mu) = X\beta \Longrightarrow g(\mu) = \eta$$

Donde la función $g(\mu)$ es conocida monótona y diferenciable de η por lo que $\mu_i = g^{-1}(\eta_i)$, tal que i varia de 1 hasta n.

Existen diferentes tipos de enlaces los cuales son: Enlaces Canónico, Enlaces Probit, Enlaces de complemento Log-log, Enlaces Logit, Enlace Box-Cox y Enlaces de Aranda-Ordaz.

A continuación se expresarán los parámetros de algunas distribuciones de la familia exponencial:

Distribución	f()	θ	ϕ
Normal: $Y \sim N(\mu, \sigma^2)$	$f(y, \theta, \phi) = exp\{\frac{1}{\sigma^2}(y\mu - \frac{2}{2}) - \frac{1}{2}(log2\pi\sigma^2 + \frac{y^2}{\sigma^2})\}$	μ	σ^2
Binomial: $Y \sim Bin(n, p)$	$f(y) = exp\{ylog(\frac{p}{1-p}) + nlog(1-p) + log\begin{bmatrix} n \\ y \end{bmatrix}\}$	$log(\frac{p}{1-p})$	n
Poisson: $Y \sim Pois(\lambda)$	$f(y) = exp\{ylog\lambda - \lambda - logy!\}$	$log\lambda$	1

Distribución	$b(\theta)$	$a(\phi)$	$c(y, \phi)$	$E(Y)$	$Var(Y)$
Normal: $Y \sim N(\mu, \sigma^2)$	$\frac{\mu^2}{2}$	ϕ	$-\frac{1}{2}[\frac{y^2}{\phi} + log(2\pi\phi)]$	μ	σ^2
Binomial: $Y \sim Bin(n, p)$	$nlog(1 - e^\theta)$	$\frac{1}{\phi}$	$\begin{bmatrix} \phi \\ \phi y \end{bmatrix}$	$\frac{1}{1+e^{-\theta}}$	$\phi^{-1}\frac{\partial\mu}{\partial\theta}$
Poisson: $Y \sim Pois(\lambda)$	e^θ	1	$logy!$	e^θ	e^θ

Modelos lineales generalizados funcionales

Los datos que observamos para el i-ésimo tema o unidad experimental son $(X_i(t), t \in \tau)$, $i = 1, ..., n$ Suponemos que estos datos forman una i.i.d. muestra. La variable predictora $X(t)$, la cual, es una curva aleatoria que se observa por sujeto o unidad experimental y corresponde a un proceso estocástico cuadrado integrable en un intervalo real τ. La variable dependiente Y es una variable aleatoria de valor real que puede ser continua o discreta. Por ejemplo, en el importante caso especial de un regresión funcional binomial, uno tendría $Y \in 0, 1$.

Suponiendo que se proporciona una función de enlace $g(\cdot)$ que es monótona y dos veces función continuamente diferenciable con derivados acotados y por lo tanto invertible. Además, tenemos una función de varianza $\sigma^2(\cdot)$ que se define en el rango de la función de enlace y es estrictamente positiva. El modelo lineal funcional generalizado o el modelo de cuasi verosimilitud funcional se determina mediante una función de parámetro $\beta(\cdot)$, que se supone que es integrable en el cuadrado en su dominio τ, además de la función de enlace $g(\cdot)$ y la función de varianza $\sigma^2(\cdot)$.

Dada una medida real dw en τ, se definen los predictores lineales

$$\eta = \alpha + \int \beta(t)X(t)dw(t)$$

y medias condicionales $\mu = g(\eta)$ donde $E(Y|X(t), t \in \tau) = \mu$ y $Var(Y|X(t), t \in \tau) = \sigma^2(\mu) = \tilde{\sigma}^2(\eta)$ para una función $\tilde{\sigma}^2(\eta) = \sigma^2(g(\eta))$. En un modelo lineal generalizado funcional, la distribución de Y se especificaría dentro de la familia exponencial. Para lo siguiente (excepto donde se indique explícitamente), será suficiente considerar el modelo de cuasi-verosimilitud funcional

$$Y_i = g(\alpha + \int \beta(t)X(t)dw(t)) + e_i \tag{1}$$

donde $i = 1, ..., n$

$$E(e|X(t), t \in \tau) = 0$$
$$Var(e|X(t), t \in \tau) = \sigma^2(\mu) = \tilde{\sigma}^2(\eta)$$

Ay que tener en cuenta que α es una constante, y la inclusión de un intercepto nos permite requiere $E(X(t)) = 0$ para toda t.

Los errores e_i son i.i.d. y usamos la integración w.r.t. de la medida $dw(t)$ para permitir funciones de peso no negativas $v(\cdot)$ tales que $v(t) > 0$ para $t \in \tau$, $v(t) = 0$ para $t \notin \tau$ y $dw(t) = v(t)dt$; la opción predeterminada será $v(t) = 1_{\{t \in \tau}$. Las funciones de peso no constantes pueden ser de interés cuando los procesos predictivos observados son estimaciones de función que pueden exhibir una mayor variabilidad en algunas regiones, por ejemplo, hacia los límites.

La función de parámetro $\beta(\cdot)$ es una cantidad de interés central en el análisis estadístico y reemplaza el vector de pendientes en un modelo lineal generalizado o estima el modelo basado en ecuaciones. Configurando $\sigma^2 = E\{\tilde{\sigma}^2(\eta)\}$, luego se encuentra

$$Var(e) = Var(e|X(t), t \in \tau) + E\{Var(e|X(t), t \in \tau)\} = E\{\tilde{\sigma}^2(\eta)\} = \sigma^2$$

Así como $E(e) = 0$.

Donde $\rho_j, j = 1, 2, ...$, es una base ortonormal del espacio funcional $L^2(dw)$, es decir, $\int_T \rho_j(t)\rho_k(t)dw(t) = \rho_{jk}$. Entonces, el proceso de predicción $X(t)$ y la función de parámetro $\beta(t)$ se pueden expandir a

$$X(t) = \sum_{j=1}^{\infty} \varepsilon_j \rho_j(t)$$

$$\beta(t) = \sum_{j=1}^{\infty} \beta_j \rho_j(t)$$

Con rv's ε_j y coeficientes β_j, dados por $\varepsilon_j = \int X(t) \times \rho_j(t)dw(t)$ y $\beta_j = \int \beta(t)\rho_j(t)dw(t)$, respectivamente.

Se tiene en cuenta que $E(\varepsilon_j) = 0$ y $\sum_j^2 = \infty$. Escribiendo $\sigma_j^2 = E(\varepsilon_j^2)$, se encuentra que $\sum \sigma_j^2 = E(X^2(t))dw(t) = \infty$. De la ortonormalidad de las funciones base ρ_j, se sigue inmediatamente

$$\int \beta(t)X(t)dw(t) = \sum_{j=1}^{\infty} \beta_j \varepsilon_j$$

Será conveniente trabajar con errores estandarizados

$$e' = e\sigma(\mu) = e\tilde{\sigma}(\eta)$$

Para lo cual $E(e'|X) = 0$, $E(e') = 0$, $E(e'^2) = 1$. Se supone que $E(e'^4) = \mu^4 < \infty$ y se nota que en el modelo (1) , la distribución de los errores e_i no necesita ser especificada y, en particular, no necesita ser un miembro de la familia exponencial. En este sentido, el modelo (1) es menos una extensión del modelo lineal generalizado clásico [*McCullagh y Nelder (1989)*] que una extensión del enfoque cuasi-verosimilitud de Wedderburn (1974). Se aborda la dificultad causada por la dimensionalidad infinita de los predictores al aproximar el modelo (1) con una serie de modelos donde el número de predictores se trunca en $p = p_n$ y la dimensión p_n aumenta asintóticamente como $n \longrightarrow \infty$.

Una motivación heurística para esta estrategia de truncamiento es la siguiente:

$$U_p = \alpha + \sum_{j=1}^{p} \beta_j \varepsilon_j$$

$$V_p = \sum_{j=p+1}^{\infty} \beta_j \varepsilon_j$$

Se encuentra que $E(Y|X(t), t \in \tau) = g(\alpha + \sum_{j=1}^{\infty} = \beta_j \varepsilon_j) = g(U_p + V_p)$. El condicionamiento en los primeros p componentes y la escritura $F_{V_p|U_p}$ para la función de distribución condicional que conduce a una función de enlace truncado g_p,

$$E(Y|U_p) = g_p(U_p) = E[g(U_p + V_p)|U_p] = \int g(U_p + s)dF_{Vp|Up}(s)$$

Para la aproximación del modelo completo por la función de enlace truncado, hay que tener en cuenta que la acotación de g', $|g'(\cdot)|^2 \leq c$, implica que

$$\left\{ \int [g(U_p + V_p) - g(U_p + s)]dF_{V_p|U_p(s)} \right\}^2$$

$$\leq \int g?(\xi)^2 (Vp?s)^2 dF_{V_p|U_p}(s)$$

$$\leq 2c \int (V_p^2 + s^2)dF_{V_p|U_p}(s)$$

y por lo tanto,

$$\begin{aligned} &E((g(U_p+V_p)-g_p(U_p))^2) \\ &\quad = E\Big(\int [g(U_p+V_p)-g(U_p+s)]dF_{V_p|U_p}(s)\Big)^2 \\ &\quad \leq 2cE(V_p^2+E(V_p^2|U_p)) = 4cE(V_p^2) \\ &\quad \leq 4c\sum_{j=p+1}^{\infty}\beta_j^2\sum_{j=p+1}^{\infty}\sigma_j^2 \end{aligned} \tag{2}$$

El error de aproximación del modelo truncado se ve directamente relacionado con $Var(V_p)$ y está controlado por la secuencia $\sigma_j^2 = Var(\varepsilon_j)$, $j = 1, 2, ...$, que para el caso especial de una base propia corresponde a una secuencia de valores propios.

Ajustando $\varepsilon_j^{(i)} = \int X_i(t)\rho_j(t)dw(t)$, el modelo completo con errores estandarizados e_i' es

$$Y_i = g\Big(\alpha+\sum_{j=1}^{\infty}\beta_j\varepsilon_j^{(i)}\Big) + e_i'\tilde{\sigma}\Big(\alpha+\sum_{j=1}^{\infty}\beta_j\varepsilon_j^{(i)}\Big), \qquad i = 1, ..., n.$$

Con predictores lineales truncados η y media μ,

$$\eta_i = \alpha+\sum_{j=1}^{p}\beta_j\varepsilon_j^{(i)}, \qquad \mu = g(\eta_i),$$

El modelo p-truncado se convierte

$$Y_i^p = g_p\Big(\alpha+\sum_{j=1}^{\infty}\beta_j\varepsilon_j^{(i)}\Big) + e_i'\tilde{\sigma_p}\Big(\alpha+\sum_{j=1}^{\infty}\beta_j\varepsilon_j^{(i)}\Big), \qquad i = 1, ..., n. \tag{3}$$

Donde $\tilde{\sigma_p}$ se define análogamente a g_p. Hay que tener en cuenta que $g(U_p) - g_p(U_p)$ y, análogamente, $\tilde{\sigma}(U_p) - \tilde{\sigma_p}(U_p)$ están limitados por el error (2). Como se supondrá que este error desaparece asintóticamente, como $p \longrightarrow \infty$, se puede en lugar de (3) trabajar con la secuencia aproximada de modelos

$$Y_i^p = g\Big(\alpha+\sum_{j=1}^{\infty}\beta_j\varepsilon_j^{(i)}\Big) + e_i'\tilde{\sigma}\Big(\alpha+\sum_{j=1}^{\infty}\beta_j\varepsilon_j^{(i)}\Big), \qquad i = 1, ..., n. \tag{4}$$

En el cual las funciones g y $\tilde{\sigma}$ son fijas. Se puede notar que las variables aleatorias $Y_i^{(p)}$ y e_i', $i = 1, ..., n$, forman matrices triangulares, $Y_{i,n}^{(pn)}$ y $e_{i,n}'$, $i = 1, ..., n$, con cambios en la distribución cuando n cambia; por simplicidad, se suprimen los índices n.

Estimación en el modelo lineal generalizado funcional

Un objetivo central es la estimación y la inferencia para la función de parámetro $\beta(\cdot)$. La inferencia para $\beta(\cdot)$ es de interés para construir regiones de confianza y probar si la función del predictor tiene alguna influencia en el resultado, en analogía a la prueba del efecto de regresión en un modelo de regresión clásico. La base ortonormal

$\{\rho_j, j = 1, 2, ...\}$ se elige comúnmente como la base de Fourier o la base formada por las funciones propias del operador de covarianza. Las funciones propias pueden estimarse a partir de los datos tal como se describe en *Rice y Silverman (1991) o Capra y M'uller (1997*). Siempre que se incluyan la estimación y la inferencia para el intercepto α, cambiamos el rango de suma para los predictores lineales η_i en el lado derecho del modelo p-truncado (3) a $\sum_0^p$ de $\sum_1^p$, estableciendo $\varepsilon_0^{(i)} = 1$ y $\beta_0 = \alpha$. A continuación, la inclusión de α en el vector de parámetros será la predeterminada.

Al fijar p por el momento, estamos en la situación del enfoque de ecuación de estimación habitual y podemos estimar el vector de parámetro desconocido $\beta^T = (\beta_0, .., \beta_p)$ resolviendo la ecuación de estimación o puntuación.

$$U(\beta) = 0 \tag{5}$$

El ajuste de la función de puntuación valorada en el vector $\varepsilon^{(i)T} = (\varepsilon_0^{(i)}, ..., \varepsilon_p^{(i)})$, $\eta_i = \sum_{j=0}^p \beta_j \varepsilon_j^{(i)}$, $\mu_i = g(\eta_i)$, $i = 1, ..., n$, se define por

$$U(\beta) = \sum_{i=1}^{n} (Y_i - \mu_i) g'(\eta_i) \varepsilon^{(i)} / \sigma^2(\mu_i) \tag{6}$$

Las soluciones de la ecuación de puntaje (5) se denotarán por

$$\hat{\beta}^T = (\hat{\beta}_0, ..., \hat{\beta}_p); \qquad \hat{\alpha} = \hat{\beta}_0 \tag{7}$$

Las matrices relevantes que juegan un rol bien conocido en la resolución de la ecuación de estimación (5) son

$$D = D_{n,p} = (g'(\eta_i)\varepsilon_k^{(i)} / \sigma(\mu_i))_{1 \le i \le n, 0 \le k \le p},$$

$$V = V_{n,p} = diag(\sigma^2(\mu_1), ..., \sigma^2(\mu_n))_{1 \le i \le n},$$

y con copias genéricas η, ε, μ de η_i, $\varepsilon^{(i)}$, μ_i, respectivamente,

$$\Gamma = \Gamma_p = (\gamma_{kl})_{0 \le k,l \le p}; \qquad \gamma_{kl} E\Big(\frac{g''^2(\eta)}{\sigma^2(\mu)} \varepsilon_k \varepsilon_l\Big), \qquad \Xi = \Gamma^{-1} = (\xi_{kl})_{0 \le k,l \le p} \tag{8}$$

Se observa que $\Gamma = \frac{1}{n} E(D^T D)$ es una matriz definida simétrica y positiva y que existe la matriz inversa Ξ. De lo contrario, se llegaría a la contradicción $E((\sum_{k=0}^p \alpha_k \varepsilon_k g(\eta)/\sigma(\mu)))^2) = 0$ para constantes distintas de cero $\alpha_0, ..., \alpha_p$.

Con los vectores $YT = (Y_1, ..., Y_n)$, $\mu^T = (\mu_1, ..., \mu_n)$, la ecuación de estimación $U(\beta) = 0$ puede reescribirse como

$$D^T = V^{1/2}(Y - \mu) = 0$$

Esta ecuación generalmente se resuelve iterativamente por el método de mínimos cuadrados ponderados iterados. Bajo nuestras suposiciones básicas, como $\frac{1}{n} E(D^T D) = \Gamma_p$ es una matriz definida positiva fija para cada p, la existencia de una solución única para cada p fijo se asegura asintóticamente.

En lo anterior se ha supuesto que se conocen tanto la función de enlace $g(\cdot)$ como la función de varianza $\sigma^2(\cdot)$. Las situaciones donde las funciones de enlace y varianza son desconocidas son comunes, y se puede extender métodos

para cubrir el caso general donde estas funciones son suaves, que para p fijo corresponde a los modelos semiparamétricos de regresión cuasi-verosimilitud considerados en *Chiou y M Üuller (1998, 1999)*. En la implementación de SPQR, se alternan los pasos de actualización no para-métricos (suavizado) y para-métrico, utilizando un modelo paramétrico razonable para el paso de inicialización. Dado que la función de enlace es arbitraria, a excepción de las restricciones de suavidad y monotonicidad, se puede exigir que las estimaciones y parámetros satisfagan $||\beta|| = 1$, $||\hat{\beta}|| = 1$ para la identificabilidad.

Para un β, $||\hat{\beta}|| = 1$, estableciendo $\hat{\eta}_i = \sum_{j=0}^{p} \hat{\beta}_j \varepsilon_j^{(i)}$, las actualizaciones de la estimación de función de enlace $g(\cdot)$ y su primera derivada $g'(\cdot)$ se obtienen suavizando (aplicando cualquier método de suavizado de dispersión razonable que permite la estimación de derivados) el diagrama de dispersión $(\hat{\eta}_i, Y_i)_{i=1,\ldots,n}$. Las actualizaciones para la estimación de la función de varianza $\sigma^2(\cdot)$ se obtienen suavizando el diagrama de dispersión $(\hat{\mu}_i, \hat{\varepsilon}_i^2)_{i=1,\ldots,n}$, donde $\hat{\mu}_i = g(\hat{\eta}_i)$ son estimaciones de respuesta media actual y $\hat{\varepsilon}_i^2) = (Y_i - \hat{\mu}_i)^2$ son residuales cuadrados actuales. El paso de actualización paramétrica continúa resolviendo la ecuación de puntuación (5), utilizando la puntuación semiparamétrica

$$U(\beta) = \sum_{i=1}^{n} (Y_i - \hat{g}(\eta_i))\hat{g'}(\eta_i)\varepsilon^{(i)}/\hat{\sigma}^2\hat{g}(\eta_i)) \tag{9}$$

Esto conduce a las soluciones $\hat{\beta}$, en analogía a (8). Para soluciones de las ecuaciones de puntuación para ambas puntuaciones (6) y (9), se obtiene las estimaciones de la función de regresión

$$\hat{\beta}(t) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j \rho_j(t) \tag{10}$$

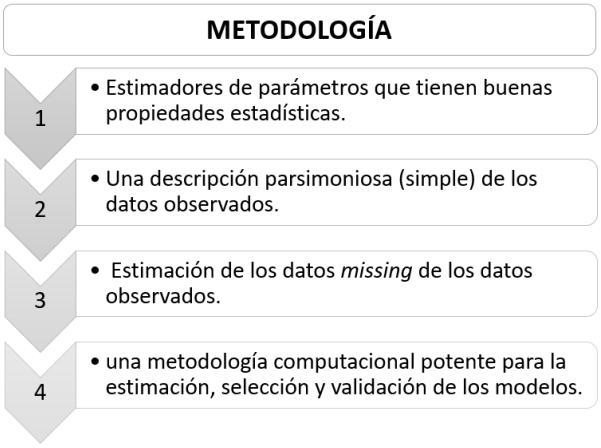
Las matrices D y Γ se modifican análogamente para el caso SPQR, sustituyendom estimaciones apropiadas.

5.2.3. Estadística bayesiana

por lo general las probabilidades se utilizan para exteriorizar la información o la incertidumbre que se tiene respecto a medida desconocida. Pero el uso de dichas probabilidades que expresan la información, se pueden elaborar de manera más formal en la estadística. De modo matemático el cálculo de probabilidades se pueden lograr desde la regla de bayes, ya que existe una relación entre probabilidad e información.

Los métodos bayesianos se derivan de los principios de inferencia bayesiana. este método produce:

- Estimadores de parámetros que tienen buenas propiedades estadísticas.
- Una descripción parsimoniosa (simple) de los datos observados.
- Estimación de los datos *missing* de los datos observados.
- una metodología computacional potente para la estimación, selección y validación de los modelos.



Diferencias entre la estadística clásica y la estadística bayesiana

Las diferencias entre la estadística clásica y la bayesiana se da a partir de la noción de probabilidad, ya que queremos tratar los parámetros desconocidos con los que representamos de forma característica el estado real. La probabilidad en la estadística clásica se da más como una valoración ideal, que se encuentra en esencia,donde los valores son fijos pero desconocidos; aquí nos basamos en la selección de los valores de los parámetros que maximizan la probabilidad de contemplar los datos.

Por otro lado, la estadística bayesiana no solo recurre a la muestra si no que también a la in formación *a priori (distribución subjetiva antes de ver los datos* que tiene una relación con los sucesos o fenómenos que se quieren modelizar.

Inferencia Bayesiana

Según *Andrew Gelman, John B. Carlin, Hal S. Stern, Davi* Las conclusiones estadísticas bayesianas sobre un parámetro θ, o datos no observados $\tilde{y}$, se hacen en términos de enunciados de probabilidad. Estas declaraciones de probabilidad están condicionadas al valor observado de y, y en nuestra notación se escriben simplemente como $p(\theta|y)$ o $p(\tilde{y}|y)$. También condicionamos implícitamente los valores conocidos de cualquier covariable, x. Es en el nivel fundamental de condicionamiento en los datos observados que la inferencia bayesiana se aparta del enfoque de inferencia estadística descrito en muchos libros de texto, que se basa en una evaluación retrospectiva del procedimiento utilizado para estimar $\theta(o\tilde{y})$ sobre la distribución de posibles y valores condicionales en el verdadero valor desconocido de θ. A pesar de esta diferencia, se verá que en muchos análisis simples, conclusiones superficialmente similares resultan de los dos enfoques de la inferencia estadística. Sin embargo, los análisis obtenidos usando métodos bayesianos pueden extenderse fácilmente a problemas más complejos.

Reglas de Bayes

Para hacer afirmaciones de probabilidad sobre θ dado y, debemos comenzar con un modelo que proporcione una distribución de probabilidad conjunta para θ y y. La función de masa o densidad de probabilidad conjunta se puede escribir como un producto de dos densidades que a menudo se denominan distribución previa $p(\theta)$ y distribución de muestreo (o distribución de datos) $p(y|\theta)$, respectivamente:

$$p(\theta, y) = p(\theta)p(y|\theta)$$

Simplemente condicionando el valor conocido de los datos y, usando la propiedad básica de probabilidad condi-

cional conocida como regla de Bayes, se obtiene la densidad posterior:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{11}$$

donde $p(y) = \sum_\theta p(\theta)p(y|\theta)$, y la suma es sobre todos los valores posibles de θ (o $p(y) = \int p(\theta)p(y|\theta)d?$ en el caso de continuo θ). Una forma equivalente de (11) omite el factor $p(y)$, que no depende de θ y, con y fijo, puede considerarse una constante, produciendo la densidad posterior no normalizada, que es el lado derecho de 12:

$$p(\theta|y) = p(\theta, y) = p(\theta)p(y|\theta) \tag{12}$$

El segundo término en esta expresión, $p(y|\theta)$, se toma aquí como una función de θ, no de y. Estas fórmulas simples encapsulan el núcleo técnico de la inferencia bayesiana: la tarea principal de cualquier aplicación específica es desarrollar el modelo (θ, y) y realizar los cálculos para resumir $p(?|y)$ de manera apropiada.

Predicción

Para hacer inferencias sobre un observable desconocido, a menudo llamado inferencias predictivas, seguimos una lógica similar. Antes de que se consideren los datos y, la distribución de lo desconocido pero observable es

$$p(y) = \int p(y, \theta)d\theta = \int p(\theta)p(y|\theta)d\theta \tag{13}$$

Esto a menudo se denomina distribución marginal de y, pero un nombre más informativo es la *distribución predictiva anterior*: anterior porque no está condicionado a una observación previa del proceso, y predictivo porque es la distribución de una cantidad que es observable. Una vez que se han observado los datos, podemos predecir un observable desconocido, $\tilde{y}$, del mismo proceso. Por ejemplo, $y = (y_1, ..., y_n)$ puede ser el vector de pesos registrados de un objeto ponderado n veces en una escala, $\theta = (\mu, \sigma^2)$ puede ser el peso verdadero desconocido del objeto y la medida varianza de la escala, y $\tilde{y}$ puede ser el peso pendiente del peso del objeto en un nuevo pesaje planeado. La distribución de $\tilde{y}$ se llama distribución predictiva posterior, posterior porque está condicionada a la observación de y es predictiva porque es una predicción para una $\tilde{y}$ observable:

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y)d\theta & (14)\\ &= \int p(\tilde{y}|\theta, y)p(\theta|y)d\theta & (15)\\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta & (16) \end{aligned}$$

La segunda y tercera líneas muestran la distribución predictiva posterior como un promedio de predicciones condicionales sobre la distribución posterior de θ. El último paso se deriva de la independencia condicional asumida de y y $\tilde{y}$ dado θ.

Modelos Lineales Generalizados

Este capítulo revisa los modelos lineales generalizados desde una perspectiva bayesiana. Discutimos distribuciones previas y modelos jerárquicos. Demostramos cómo aproximar la probabilidad mediante una distribución normal, como una aproximación y como un paso en cómputos más generales. Finalmente, discutimos la clase de modelos loglineales, una subclase de modelos lineales generalizados de Poisson que se usa comúnmente para la

imputación de datos faltantes y resultados discretos multivariantes. Este capítulo no pretende ser exhaustivo, sino más bien proporcionar orientación suficiente para que el lector pueda combinar modelos lineales generalizados con las ideas de modelos jerárquicos, simulación posterior, predicción, verificación de modelos y análisis de sensibilidad que ya hemos presentado para métodos bayesianos en general, y modelos lineales en particular.

Los modelos lineales generalizados unifica los enfoques necesarios para analizar los datos para los cuales no es apropiada la suposición de una relación lineal entre x e y o la suposición de variación normal. Un modelo lineal generalizado se especifica en tres etapas:

1. El predictor lineal, $\eta = X\beta$,
2. La función de enlace $g(\cdot)$ que relaciona el predictor lineal con la media del resultado variable: $\mu = g^{-1}(\eta) = g^{-1}(X\beta)$,
3. El componente aleatorio que especifica la distribución de la variable de resultado y con la media $E(y|X) = \mu$. La distribución también puede depender de un parámetro de dispersión, ϕ.

Apartir de Aquí el análisis va a estar enfocado a la distribución binomial

---

Modelos lineales generalizados estandarizados

Continuos

La regresión lineal es un caso especial del modelo lineal generalizado, con la función de enlace de identidad, $g(\mu) = \mu$. Para los datos continuos que son todos positivos, podemos usar el modelo normal en la escala logarítmica. Cuando esa familia de distribución no se ajusta a los datos, las distribuciones de gamma y Weibull a veces se consideran alternativas.

Binomial

Quizás el más utilizado de los modelos lineales generalizados es el de los binarios o datos binomiales. Supongamos que y_i $Bin(n_i, \mu_i)$ con n_i conocido. Es común especificar el modelo en términos de la media de las proporciones y_i/n_i, en lugar de la media de y_i. Elegir la transformación logit de la probabilidad de éxito, $g(\mu_i) = log(\mu_i/(1-\mu_i))$, como la unción de enlace conduce al modelo de regresión logística. La distribución para los datos y es

$$p(y|\beta) = \Pi_{i=1}^{n}\Big(\frac{n_i}{y_i}\Big)\Big(\frac{e^{\eta_i}}{1+e^{\eta_i}}\Big)^{y_i}\Big(\frac{e^{\eta_i}}{1+e^{\eta_i}}\Big)^{n_i-y_i}$$

Distribuciones previas no informativas sobre β_i

El análisis clásico de modelos lineales generalizados se obtiene si se supone una distribución previa no informativa o plana para β. El modo posterior correspondiente a una densidad previa uniforme no informativa es la estimación de máxima verosimilitud para el parámetro β, que puede obtenerse usando regresión lineal ponderada iterativa (como se implementa en los paquetes informáticos R o Glim, por ejemplo). La inferencia posterior aproximada se puede obtener a partir de una aproximación normal a la probabilidad.

Ejemplo. El Modelo Logístico Binomial

En el modelo lineal generalizado binomial con enlace logístico, la probabilidad logarítmica para cada observación es

$$L(y_i|\eta_i) = y_i log\Big(\frac{e^{\eta_i}}{1+e^{\eta_i}}\Big) + (n_i - y_i) log\Big(\frac{e^{\eta_i}}{1+e^{\eta_i}}\Big) \tag{17}$$

$$= y_i\eta_i - n_i log(1+e^{\eta_i}). \tag{18}$$

Según la *Facultad de Ingeniería de la la Universidad de la República de Uruguay* La distribución a priori cumple un papel importante en el análisis bayesiano ya que mide el grado de conocimiento inicial que se tiene de los parámetros en estudio. Si bien su influencia disminuye a medida que más información muestral es disponible, el uso de una u otra distribución a priori determinara ciertas diferencias en la distribución a posteriori.

Si se tiene un conocimiento previo sobre los parámetros, este se traducirá en una distribución a priori. Así,será posible plantear tantas distribuciones a priori como estados iniciales de conocimiento existan y los diferentes resultados obtenidos en la distribución a posteriori bajo cada uno de los enfoques, adquirirán una importancia en relación con la convicción que tenga el investigador sobre cada estado inicial. Sin embargo, cuando nada es conocido sobre los parámetros, la selección de una distribución a priori adecuada adquiere una connotación especial pues será necesario elegir una distribución a priori que no influya sobre ninguno de los posibles valores de los parámetros en cuestión. Estas distribuciones a priori reciben el nombre de difusas o no informativas y en esta sección se tratara algunos criterios para su selección.

$$\frac{dL}{d\eta_i} = y_i - n_i \frac{e^{\eta_i}}{1+e^{\eta_i}} \tag{19}$$

$$\frac{d^2L}{d\eta_i^2} = -n_i \frac{e^{\eta_i}}{(1+e^{\eta_i})^2} \tag{20}$$

Por lo tanto, los pseudo datos z_i y la varianza σ_i^2 de la aproximación normal para la i-ésima unidad de muestreo son

$$z_i = \hat{\eta}_i + \frac{(1+e^{\hat{\eta}_i})^2}{e^{\hat{\eta}_i}}\Big(\frac{y_i}{n_i}\frac{e^{\hat{\eta}_i}}{1+e^{\hat{\eta}_i}}$$

$$\sigma_i^2 = \frac{1}{n_i}\frac{(1+e^{\hat{\eta}_i})^2}{e^{\hat{\eta}_i}}.$$

La aproximación depende de $\hat{\beta}$ a través del predictor lineal $\hat{\eta}$.

El modo posterior se puede encontrar utilizando la regresión lineal ponderada iterativa: en cada paso, uno calcula la aproximación normal a la probabilidad en función de la conjetura actual de (β, ϕ) y encuentra el modo de la distribución posterior aproximada resultante mediante regresión lineal ponderada. (Si hay información previa disponible en β, debe incluirse como filas adicionales de datos y variables explicativas en la regresión para información previa fija. Iterar este proceso equivale a resolver el sistema de k ecuaciones no lineales, $\frac{dp(\beta|y)}{d\beta}) = 0$, utilizando el método de Newton, y converge al modo rápidamente para los modelos lineales generalizados estándar. Una posible dificultad son las estimaciones de β que tiende al infinito, lo que puede ocurrir, por ejemplo, en la regresión logística si hay algunas combinaciones de variables explicativas para las cuales μ es casi igual a cero o uno. La información previa sustancial tiende a eliminar este problema.

Si un parámetro de dispersión, ϕ, está presente, se puede actualizar ϕ en cada paso de la iteración maximizando su densidad posterior condicional (que es unidimensional), dada la conjetura actual de β. De manera similar, uno puede incluir en la iteración cualquier parámetro de varianza jerárquica que necesite ser estimado y actualizar sus valores en cada paso.

Distribución posterior normal aproximada

Una vez que se ha alcanzado el modo $(\hat{\beta}, \hat{\phi})$, se puede aproximar la distribución posterior condicional de β, dada $\hat{\phi}$, por el resultado del cómputo de regresión lineal ponderado más reciente; es decir,

$$p(\beta|\hat{\phi}, y) \approx N(\beta|\hat{\beta}, V\beta),$$

donde V_β en este caso es $(X^T diag(-L''(y_i, \eta\eta\eta_i, \phi))X)^{-1}$. (En general, solo se necesita calcular el factor de Cholesky de V_β. Si el tamaño de muestra n es grande, y ϕ no es parte del modelo (como en las distribuciones binomial y de Poisson), podemos estar contenido para detener aquí y resumir la distribución posterior por la aproximación normal a $p(\beta|y)$. Si un parámetro de dispersión, ϕ, está presente, uno puede aproximar la distribución marginal de ϕ, así (No hay un parámetro de dispersión ϕ en el modelo binomial.)

$$p_{approx}(\phi|y) = p(\beta, \phi|y)p_{approx}(\beta|\phi, y) \approx p(\hat{\beta}(\phi), \phi|y)|V_\beta(\phi)|^{1/2},$$

Donde β y V_β en la última expresión son el modo y la matriz de varianza de la aproximación normal condicional en ϕ.

6. Metodología

Tipo de estudio

El presente trabajo es de carácter estimativo, predictivo y de visualización, para datos funcionales, donde los datos son procesos estocástico por país, lo que permite encontrar el comportamiento funcional.

Método

Al inicio del documento se ajusta toda la parte teórica, de los índices que componen la base de datos, además se da lugar a la explicación teórica del análisis de datos funcionales, tal como análisis descriptivo, exploratorio, multivariado y modelos. luego se continuara con la explicación de los modelos generalizado para posteriormente explicar toda la parte bayesiana.

Dentro de los modelos lineales generalizados el modelado estadístico son herramientas metodológicas que permiten codificar todas las situaciones de análisis dentro de un mismo esquema general. Aquí se explicara la familia exponencial y también dicho modelo de manera funcional, haciendo énfasis a la distribución binomial.

En la teoría bayesiana se realiza un enfoque al modelo lineal generalizado binomial con prior no informativa, ya que con tal prior se realizaran las estimaciones.

Terminada toda la parte teórica se hará la implementación del análisis a los datos escogidos, tal como:

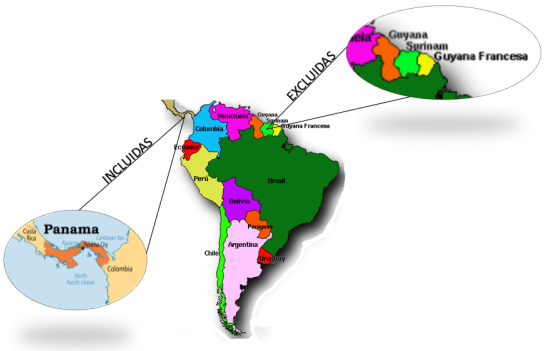
1. Organización de la base para un eficaz análisis.
2. Identificación del tipo de datos funcionales.
3. Visualización del comportamiento de cada índice por país.
4. Suavizamiento de las funciones (si es necesario) y análisis descriptivo.
5. Detección de países atípicos de América del sur.

6. Estimación clásica del modelo lineal generalizado binomial.

7. Estimación bayesiana del modelo lineal generalizado binomial, por medio del algoritmo seleccionado.

7. Resultados

En esta sección se muestra el análisis de los datos, los cuales están compuestos por: (1)[5] 11 países de América del Sur , (2)[6] 4 índices económicos (3)[7] desde el año 2002 hasta el 2015.



7.1. Organización de la base para un eficaz análisis

Para realizar los posterires análisis en el programa R, la base de datos se organiza de la siguiente manera[8]:

La base de datos es de dimención $14x45$ [9]

[5] Son 10 países propios de América del Sur que se analizan, pero se incluye Panamá, ya que espacial-mente y culturalmente se encuentran más relacionados a dicho continente; además se excluye Surinam y Guyana, por que el Banco Mundial no realiza publicaciones del índice de GINI en ese lapso de tiempo y el Índice EIU desde el 2000 hasta el 2005

[6] índice EIU, PIB Per Cápita y índice de GINI

[7] Los datos se trabajan desde el año 2002, ya que en el 2001 el banco mundial no realizó la publicación del índice EIU; tampoco se trabajo el 2016, debido a las alteraciones en la política y la economía que se presento en Venezuela, en el gobierno de Nicolás Maduro

[8] ARG=Argentina, BOL=Bolivia, BRA=Brasil, CHL=Chile, COL=Colombia, ECU=Ecuador, PAN=Panama, PRY=Paraguay, PER=Perú, URY=Uruguay, VEN=Venezuela

[9] Siendo el **Índice EIU** desde la columna 2 hasta la 12, el **PIB Per Capita** desde la columna 13 hasta la 23 y el **Índice de GINI** desde la columna 24 hasta la 34.

AÑOS	ARG	BOL	BRA	CHL	COL	ECU	PAN	PRY	PER	URY	VEN	...
2002	0	0	0	1	0	0	0	0	0	1	0	...
2003	0	0	0	1	0	0	0	0	0	1	0	...
2004	0	0	0	1	0	0	0	0	0	1	0	...
2005	0	0	0	1	0	0	0	0	0	1	0	...
2006	1	0	1	1	1	0	1	1	1	1	0	...
2007	1	0	0	1	0	0	1	0	0	1	0	...
2008	1	1	1	1	1	0	1	1	1	1	0	...
2009	1	0	0	1	0	0	1	0	0	1	0	...
2010	1	0	1	1	1	0	1	1	1	1	0	...
2011	1	0	1	1	1	0	1	1	1	1	0	...
2012	1	0	1	1	1	0	1	1	1	1	0	...
2013	1	0	1	1	1	0	1	1	1	1	0	...
2014	0	0	0	1	0	0	1	0	0	1	0	...
2015	1	0	0	1	0	0	1	0	0	1	0	...

...	ARG	BOL	BRA	CHL	COL	ECU	PAN	PRY	PER	...
...	2579, 19	913, 58	2819, 65	4463, 55	2355, 73	2183, 97	4126, 14	2059, 19	1148, 23	...
...	3330, 44	917, 36	3059, 59	4787, 7	2246, 26	2440, 47	4267, 14	2180, 25	1174, 78	...
...	4251, 57	978, 33	3623, 05	6210, 83	2740, 25	2708, 56	4591, 89	2448, 14	1408, 53	...
...	5076, 88	1046, 43	4770, 18	7615, 3	3386, 03	3021, 94	4916, 55	2754, 78	1507, 15	...
...	5878, 76	1233, 59	5860, 15	9484, 68	3709, 08	3350, 79	5348, 52	3171, 5	1809, 71	...
...	7193, 62	1389, 63	7313, 56	10526, 88	4674, 22	3590, 72	6068, 09	3611, 21	2312, 19	...
...	8953, 36	1736, 94	8787, 61	10781, 37	5433, 72	4274, 95	6973, 93	4208, 88	3059, 99	...
...	8161, 31	1776, 87	8553, 38	10243, 33	5148, 42	4255, 57	7429, 63	4166, 09	2599, 6	...
...	10276, 26	1981, 16	11224, 15	12860, 18	6250, 66	4657, 3	7937, 26	5022, 49	3225, 59	...
...	12726, 91	2377, 68	13167, 47	14705, 69	7227, 74	5223, 35	9270, 72	5771, 57	3988, 01	...
...	12969, 71	2645, 23	12291, 47	15431, 9	7884, 98	5702, 1	10589, 83	6387, 79	3855, 54	...
...	12976, 64	2947, 94	12216, 9	15941, 4	8030, 59	6074, 09	11685, 98	6583, 12	4479, 91	...
...	12245, 26	3124	12026, 62	14817, 38	7913, 38	6432, 22	12593, 74	6491, 05	4712, 82	...
...	13467, 1	3077, 03	8757, 21	13653, 23	6044, 53	6205, 06	13134, 04	6030, 34	4109, 37	...

...	URY	VEN	ARG	BOL	BRA	CHL	COL	ECU	PAN	...
...	4088, 77	3655, 98	53, 79	60, 16	58, 62	53, 3	58, 25	56	56, 59	...
...	3622, 05	3232, 52	53, 54	63, 15	58, 01	54, 6	54, 41	54, 99	56, 37	...
...	4117, 31	4271, 37	50, 18	55, 01	56, 88	52	56, 11	54, 12	55, 06	...
...	5220, 95	5432, 69	49, 27	58, 47	56, 64	51	55, 04	54, 12	53, 99	...
...	5877, 88	6735, 8	48, 26	56, 87	55, 93	52, 2	60, 8	53, 2	55, 06	...
...	7009, 7	8318, 8	47, 37	55, 44	55, 23	52, 5	59, 37	54, 33	52, 97	...
...	9062, 31	11227, 23	46, 27	51, 43	54, 37	50, 6	56, 04	50, 61	52, 63	...
...	9415, 17	11536, 15	45, 27	49, 65	53, 87	48, 7	55, 92	49, 28	52, 03	...
...	11938, 21	13545, 21	45, 3	49, 7	53, 9	51, 2	55, 9	49, 3	51, 9	...
...	14166, 5	10741, 58	43, 57	46, 26	53, 09	50, 84	54, 18	46, 21	51, 83	...
...	15092, 07	12755	42, 49	46, 7	52, 67	52, 1	53, 54	46, 57	51, 9	...
...	16881, 21	12237, 22	42, 28	48, 06	52, 87	50, 45	53, 49	47, 29	51, 66	...
...	16737, 9	6772	42, 67	48, 4	51, 48	50	53, 5	45, 38	50, 7	...
...	15524, 84	4263	38, 56	47	51, 7	48, 2	52, 5	47, 6	50, 2	...

...	PRY	PER	URY	VEN	ARG	BOL	BRA	CHL	COL	...
...	54, 04	57, 34	46, 66	50, 56	63, 91	50, 73	59, 6	68, 69	52, 06	...
...	53, 71	55, 55	46, 22	50, 37	62, 57	51, 83	58, 42	69, 65	53, 03	...
...	51, 2	52, 59	47, 13	49, 82	62, 37	53, 46	59, 45	71, 64	53, 43	...
...	51, 84	51, 37	45, 87	52, 36	60, 3	54, 38	59, 94	72, 04	52, 54	...
...	51, 67	53, 63	47, 2	46, 94	60, 35	54, 42	59, 82	73, 19	57, 24	...
...	51, 35	52, 09	47, 63	42, 2	60, 36	53, 96	60, 61	74, 31	57, 5	...
...	48, 55	51, 04	46, 27	41, 1	59, 95	53, 82	59, 38	74, 18	56, 48	...
...	47, 96	49, 67	46, 28	41	59, 42	53, 73	59, 83	73, 24	56, 89	...
...	51, 8	46, 2	45, 3	39, 4	59, 52	52, 97	61, 44	73, 19	57, 42	...
...	45, 48	52, 6	43, 37	39, 7	58, 31	52, 76	61, 01	72, 5	59, 7	...
...	45, 11	48, 17	41, 32	40, 5	57, 89	53, 73	60, 95	71, 59	59, 25	...
...	44, 73	48, 3	41, 87	40, 7	57, 8	55, 31	60, 51	71, 11	60, 15	...
...	44, 14	51, 67	41, 6	39, 8	58, 54	53, 99	61, 4	72, 23	60, 34	...
...	49, 3	44, 7	42, 6	40, 5	92, 83	76, 72	94, 23	77, 94	80, 38	...

...	PRY	PER	URY	VEN
...	54, 04	57, 34	46, 66	50, 56
...	53, 71	55, 55	46, 22	50, 37
...	51, 2	52, 59	47, 13	49, 82
...	51, 84	51, 37	45, 87	52, 36
...	51, 67	53, 63	47, 2	46, 94
...	51, 35	52, 09	47, 63	42, 2
...	48, 55	51, 04	46, 27	41, 1
...	47, 96	49, 67	46, 28	41
...	51, 8	46, 2	45, 3	39, 4
...	45, 48	52, 6	43, 37	39, 7
...	45, 11	48, 17	41, 32	40, 5
...	44, 73	48, 3	41, 87	40, 7
...	44, 14	51, 67	41, 6	39, 8
...	49, 3	44, 7	42, 6	40, 5

Cabe resaltar que inicialmente los datos fueron tomados del **Banco Mundial**, pero en algunos años que no habían datos, completaron con publicaciones de paginas oficiales de economía del respectivo país.

7.2. Identificación del tipo de datos funcionales

Base B-Spline

Los objetos de datos funcionales se construyen especificando un conjunto de funciones básicas y un conjunto de coeficientes que definen una combinación lineal de estas funciones básicas. La base B-spline se usa para funciones no periódicas. Las funciones de base B-spline son segmentos polinomiales unidos de extremo a extremo en valores de argumento llamados nudos, rupturas o puntos de unión. Los segmentos tienen una suavidad especificable a través de estos descansos. Las funciones de base B-splne tienen las ventajas de un cálculo muy rápido y una gran flexibilidad. Una base poligonal generada por **create.polygonal.basises** esencialmente una base B-spline de orden 2, grado 1. Las bases monomiales y polinómicas se pueden obtener como transformaciones lineales de ciertas bases B-spline. `http://bit.ly/2j1vXZ2` R Documentation

7.3. Visualización del comportamiento de cada índice por país

7.3.1. índice EIU

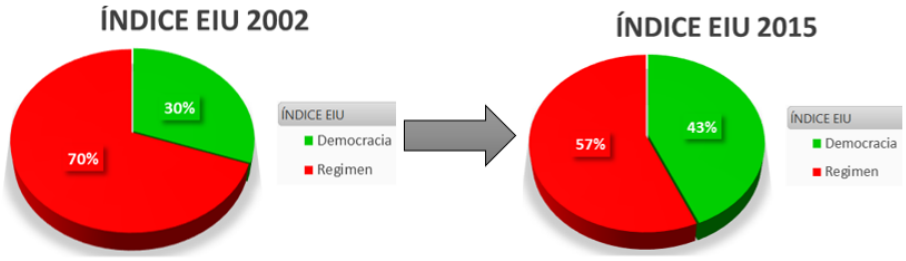


Figura 1: Comparación del índice EIU entre 2002 y 2015 en América del sur

En la Figura 1 se observa que en el 2002 el 70 % de América del Sur se encuentra bajo un estado de Régimen, lo que quiere decir que se encuentra en pluralismo político. Muchos países de esta categoría son dictaduras absolutas y las elecciones tienen irregularidades sustanciales que a menudo previenen que sean libres y justas. La presión del gobierno sobre los partidos de la oposición y los candidatos puede ser común. También es importante resaltar que el *Régimen* existe represión de las críticas hacia el gobierno y hay censura generalizada. No existe un poder judicial.

PAIS	COD_PAIS	2002	2003	2004	2005	2006	2007	2008
Argentina	ARG	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia	Democracia
Bolivia	BOL	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia
Brasil	BRA	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia
Chile	CHL	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia
Colombia	COL	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia
Ecuador	ECU	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen
Panama	PAN	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia	Democracia
Paraguay	PRY	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia
Perú	PER	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Democracia
Uruguay	URY	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia
Venezuela	VEN	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen

PAIS	COD_PAIS	2009	2010	2011	2012	2013	2014	2015
Argentina	ARG	Democracia	Democracia	Democracia	Democracia	Democracia	Régimen	Democracia
Bolivia	BOL	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen
Brasil	BRA	Régimen	Democracia	Democracia	Democracia	Democracia	Régimen	Régimen
Chile	CHL	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia
Colombia	COL	Régimen	Democracia	Democracia	Democracia	Democracia	Régimen	Régimen
Ecuador	ECU	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen
Panama	PAN	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia
Paraguay	PRY	Régimen	Democracia	Democracia	Democracia	Democracia	Régimen	Régimen
Perú	PER	Régimen	Democracia	Democracia	Democracia	Democracia	Régimen	Régimen
Uruguay	URY	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia	Democracia
Venezuela	VEN	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen	Régimen

Las Naciones Unidas tienen una lista de territorios no auto-gobernados, que fue inicialmente preparada en 1946 de acuerdo con el artículo XI de la Carta de las Naciones Unidas. Esta lista es actualizada por la Asamblea General por recomendación del Comité Especial de Descolonización; en el caso de América del sur no hay ningún país que tenga territorios no auto-gobernados Los territorios dependientes no se han separado de los países que los administran, aunque muchos de ellos tengan un alto grado de auto-gobierno y organizaciones políticas independientes como lo son: Sector Antártico Argentino de Argentina y el Territorio Chileno Antártico de Chile , lo cual también tiene influencia en el calculo del índice.

En el 2008 las libertades de los medios se erosionaron en toda América Latina y las fuerzas populistas con dudosas credenciales democráticas pasaron a primer plano. http://bit.ly/2i1DqGV The Economist Intelligence Unit's Index of Democracy 2008.

Según la Universidad de Costa Rica en una publicación realizada en el 2016, en 1990 el sistema judicial costarricense experimentó una ola modernizadora cuyo objetivo era la capacitación para atacar la creciente sofisticación de los delitos de corrupción. A 2015, este sistema se considera uno de los menos corruptos de Latinoamérica, pero sigue teniendo importantes debilidades, dentro de las cuales la percepción de falta de protección para los informantes. http://bit.ly/2Bjzadq Índices de Democracia, Corrupción y Paz.

7.3.2. PIB Per Cápita

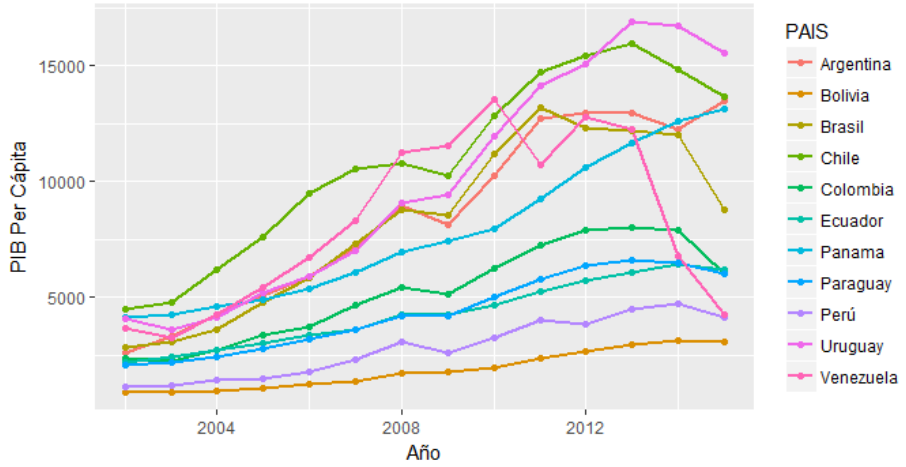


Figura 2: Comportamiento desde el 2002 hasta el 2015 del PIB Per cápita

Venezuela

A partir del 2012 el PIB Per Cápita comienza a decaer de manera acelerada, esto se debe a los acontecimientos políticos y económicos que se han venido presentado. En los últimos cuatro años la producción per cápita de Venezuela cayó $27,7\,\%$. Solo el año pasado la disminución fue de $17,9\,\%$, señaló Francisco Rodríguez, economista jefe de Torino Capital LLC, en su informe: Venezuela esta semana: Sigue y sigue. Precisó que hubo dos períodos en el país en los que el producto interno bruto descendió $23,3\,\%$ y $24\,\%$. El primero fue en los años ochenta con la crisis de la deuda y la caída de los precios del petróleo. El segundo se relaciona también con un declive histórico de precios del petróleo y los conflictos políticos que marcaron los primeros años de gobierno de Hugo Chávez. Sin embargo, estos descensos se produjeron durante períodos de seis a ocho años. Mientras que la caída asociada con Nicolás Maduro en la presidencia ha ocurrido en cuatro años, aseveró. (*NACIONAL, E. (2017). Producción per cápita de Venezuela cayó* $27,7\,\%$ *en 4 años. El Nacional*). http://bit.ly/2iHGg1x

Sin embargo, las causas de ese deterioro en el poder adquisitivo hay que buscarlas mucho antes de Maduro. Los problemas comenzaron desde que Hugo Chávez llegó al poder e inició su campaña de "nacionalizaciones, proteccionismo, regulación excesiva y obstáculos a la existencia de un sistema de precios funcional", según el estudio. (*Infobae. (2017). El ingreso per cápita de los venezolanos disminuyó* $28\,\%$ *en cuatro años.*) http://bit.ly/2ka8Rmn

Según Antonella Marty en PANAMPOST(Noticias y análisis de las américas), 2017. En 1980, Venezuela tenía el mayor PIB per cápita de América Latina, siendo uno de los países más desarrollados de la región. Pero, ¿qué sucedió?: llegó el socialismo.

Gracias a los diferentes medios de comunicación se conoce la situación de hambre, pobreza, violencia, inseguridad, inflación, escasez e involución que abunda en Venezuela. No obstante, veamos esta realidad en cifras:

1. En enero de 2012, Venezuela llegó al primer lugar del índice Big Mac de The Economist, teniendo la hamburguesa más cara del mundo ($US\$9,08$). Dan igual las variaciones, hoy ya ni siquiera hay hamburguesas, no hay carne y tampoco pan, pero cuidado, hay patria y revolución.
2. En 2007 el salario mínimo era de 466 bolívares (entonces USD $\$113,66$), la canasta básica de 2.054 bolívares (entonces USD $\$500,98$), por lo que se necesitaban 4,4 salarios. En 2017 el salario mínimo 40.638 bolívares (US $\$7,81$), y la canasta básica de 832.259,95 bolívares (USD $\$159,99$), por lo que se necesitan 20,5 salarios. En Cuba los trabajadores ganan USD $\$27,92$, es decir 0,93 al día, siendo este el segundo salario promedio más bajo de la región. Tal y como indica El Estímulo, los maestros en Venezuela ganan USD $\$33,86$ al mes, mientras que en Colombia un profesor de enseñanza primaria con tres años de experiencia gana aproximadamente USD $\$1.300$ mensuales. En Argentina alrededor de USD $\$1.370$ y en Chile USD $\$1.600$, de acuerdo con la información encontrada en Elsalario.com.
3. Según Chelminski, tan mal se ha manejado el país y tanto se han empobrecido los venezolanos, que si bien Bs. 4,30 en enero de 1983 compraban USD $\$1$, esos mismos Bs. 4,30 al cierre de 2016, solo compraban USD $\$0,00000179$ (entiéndase 179 cien millonésimas de un dólar).
4. Venezuela tiene la inflación más alta del mundo, estimada por el Fondo Monetario Internacional con un aumento para este año, llegando hasta $1.640\,\%$ durante el 2017.
5. En los últimos 15 años, más de $1.600.000$ venezolanos han emigrado según un estudio de la Universidad Simón Bolivar, es decir, casi un $5\,\%$ de la población de dicha nación.
6. , Solo el $4,8\,\%$ de los ciudadanos venezolanos tiene una percepción positiva de la situación del país.
7. , Más de 110 toneladas métricas de cocaína pasan por Venezuela cada año. Según la Asuntos Narcóticos de los Estados Unidos de América, más de la mitad de la droga colombiana pasa primero por Venezuela.
8. Según el informe de drogas del Departamento de Estado de los Estados Unidos de América, las autoridades venezolanas no persiguen eficazmente al narcotráfico. Claro, porque las autoridades venezolanas del régimen son el narcotráfico.
9. Venezuela es territorio seguro para los narcos. Durante el régimen de Hugo Chávez, más de quince capos y cabecillas del narcotráfico mundial se refugiaron en Venezuela para desarrollar sus negocios ilegales de droga, operando bajo la protección de la Fuerza Armada Nacional Bolivariana y funcionarios del régimen. Fue Chávez quien ordenó el cese de actividades en 2005 de la Administración para el Control de Drogas (DEA por sus siglas en Inglés).
10. Entre 2004 y 2013, el Tribunal Supremo de Justicia no dictó ni una sola sentencia en contra de los narcodictadores del régimen.

(*PanAm Post. (2017). La ruina socialista de Venezuela en cifras: del mayor PIB per cápita de América Latina al país más miserable del mundo.*)

http://bit.ly/2ideYpH

Bolivia

Como se puede observar en la Figura 2, Bolivia en todos los años a tenido menor valor en toda América del sur, lo que quiere decir que sus habitantes tienen un bajísimo nivel de vida. En la pagina El Día el economista Armando

Méndez y expresidente del Banco Central de Bolivia (BCB), remarcó que Bolivia sigue siendo uno de los últimos países de la región, pese a que hubo mejoras significativas en los últimos diez años. "Lo que vaya a pasar en los próximos años de seguir con el crecimiento del PIB per cápita, dependerá necesariamente de cuánta inversión extranjera ingrese al país. "Sin esa premisa, será difícil avisorar su crecimiento y si posible ir achicando la brecha. En este caso, el gobierno hace bien en buscar inversiones fuera del país y por otro, con los créditos chinos (por más de $\$us 7.000$ millones) que podrían hacer que el país no vaya a perder su dinamismo". apuntó. En esa línea, Julio Alvarado describe que el PIB per cápita, si bien se ha triplicado en los últimos diez años, es apenas un promedio de lo que gana un ciudadano boliviano, en esa medida el PIB no necesariamente refleja el nivel de vida de la población de manera equitativa. .[Es]a riqueza que se ha multiplicado por tres no se ha distribuido equitativamente. Podemos citar como ejemplo el caso de Potosí, donde los cooperativistas mineros en los últimos años han amasado fortunas, pero no todos los potosinos han compartido esos ingresos. Las brechas entre unos y otros siguen siendo los mismos e iguales que años anteriores", señaló. A su vez, el economista de la Fundación Milenio, Roberto Laserna, en su análisis si bien destaca el crecimiento del PIB per cápita de Bolivia como significativa, aún sigue pendiente de que el mismo sea sostenible en el tiempo, dada la actual coyuntura económica del mundo agravado por una crisis y caída de precios de las comoditties. En una situación didáctica. Según Molina, el PIB per cápita es un indicador económico en el que cada ciudadano tendría en su poder un determinado monto de recursos, pero en la realidad concretamente no necesariamente sucede porque algunos perciben montos elevados, medios, mínimos y nulos, la situación en nuestro país es elocuente. Por ejemplo, en los países desarrollados, emergentes y latinoamericanos existen personas que no cuentan con un ingreso que les permita cubrir sus necesidades básicas de alimentación, salud, educación y vivienda. Pero también hay lo otro, personas que acumularon riqueza como por ejemplo Bill Gates (norteamericano) y Carlos Slim (mexicano), cuyas fortunas alcanzan a todo lo que produce en un año nuestro país. .[En]tonces es un indicador útil pero tiene sus limitaciones para realizar análisis", sintetiza.

(*El Día. (2015). Con PIB per cápita de* $\$us 3.235$ *Bolivia es penúltimo*)

`https://www.eldia.com.bo/index.php?cat=1&pla=3&id_articulo=185819`

7.3.3. Índice de GINI

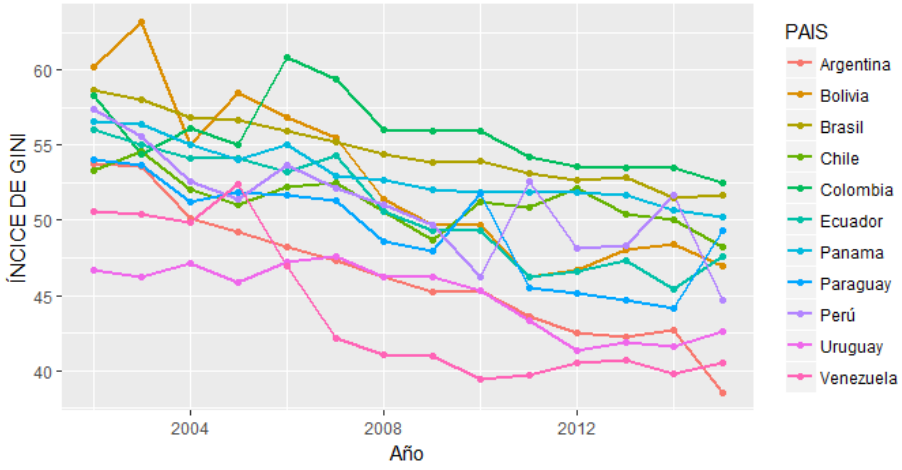


Figura 3: Comportamiento desde el 2002 hasta el 2015 del Índice de GINI

Bolivia

En el transcurso del tiempo estudiado, Bolivia es el país que llega a presentar mayor desigualdad de ingresos,

este acontecimiento ha venido afectando desde antes del 2002. En dicho país la pobreza aumentó entre 1999 y 2002 en 2.57 puntos porcentuales, y la pobreza extrema aumento 0.93 puntos porcentuales destacándose el hecho de que existe una menor proporción de personas indigentes en comparación del 2000 y 2001. Este suceso se debe al deterioro de los ingresos, de quienes se encontraban trabajando en las actividades económicas que fueron influenciadas con mayor intensidad por los *shocks* internos y externos; las actividades que se vieron más afectadas, es decir que presentaron los niveles más altos en reducción de su producción y por ende ingresos de la población ocupada son el sector Agropecuario y Transporte. (*Anon, (2017). Pobreza y Distribución del Ingreso en Bolivia entre 1999 y 2002.*) file:///C:/Users/USER/Downloads/Pobreza-99-2002.pdf

Es importante destacar que, la reducción de la desigualdad en Bolivia se produjo con mayor rapidez que en otros países de Latinoamérica, es así que de ser el segundo país con mayor desigualdad en 2005, pasó a estar entre los países con menor desigualdad en 2012. (*Medios.economiayfinanzas.gob.bo. (2014). Cite a Website - Cite This For Me.*)

http://bit.ly/2AiQb89

En el año 2003 el índice de GINI resulta ser más alto que en los demás años, Jiménez y Lizárraga (2003), muestra que los ingresos en el área rural de Bolivia están fuertemente concentrados, ya que obtienen un índice de Gini igual a 0.61 para el año 2002. Este estudio descompone el índice de Gini según las fuentes de ingreso utilizando la metodología de Leibrant et al. (1996). Los resultados muestran que la distribución de ingresos no agropecuarios contribuye al 42 por ciento de la desigualdad total de ingreso familiar. (*Villegas, H. (2006). Desigualdad en el área rural de Bolivia: ¿cuán importante es la educación?. Scielo.org.bo.*)

http://bit.ly/2zOWlQk

Venezuela

Gráficamente se puede observar que Venezuela ha tenido menor desigualdad de ingresos, tal como lo destacó el vicepresidente venezolano para el Área Social, Héctor Rodríguez, que su país es el menos desigual de América Latina, de acuerdo con su posición en el índice de GINI, método utilizado para medir la desigualdad en la distribución de los ingresos, y gracias al proyecto de nación que ideó el líder de la Revolución Bolivariana, Hugo Chávez.

Con un índice de GINI de 0,390, el cual oscila entre 0 para el país con mayor igualdad y 1 para el de más desigualdad, Venezuela es un país cada vez menos desigual, precisó Rodríguez, durante su alocución en el Seminario Suramericano sobre Inclusión Social que se realizó este jueves en Caracas.

En la cita participaron representantes de la Unión de Naciones Suramericana (Unasur), a quienes Rodríguez recordó que antes de la llegada del Comandante Hugo Chávez al poder, en el año 1998, el índice GINI se ubicaba en 0,486.

En ese sentido, el también ministro de Educación afirmó que el indicador que ubica a Venezuela más cerca de la igualdad social es consecuencia directa de la redistribución de la riqueza y del poder al pueblo, uno de los principios fundamentales del Gobierno Bolivariano.

Çhávez creó un sistema de redistribución de riqueza, de redistribución del poder para el pueblo, impulsando que sea el mismo pueblo quien se apropie del poder para trasformar su realidad, la de su comunidad, e incidir en la transformación de la vida de la Patria", explicó el funcionario.

Además, destacó el éxito de las políticas sociales implementadas por la Revolución, dirigidas a masificar la educación y los planes alimentarios, principalmente, hacia los sectores populares, así como una inversión social de 623 mil 580 millones de dólares en los últimos 15 años. (*Telesurtv.net. (2014). Venezuela es la nación menos desigual de América Latina.*)

http://bit.ly/2Bv7cMs

7.4. Suavizamiento de las funciones (si es necesario) y análisis descriptivo

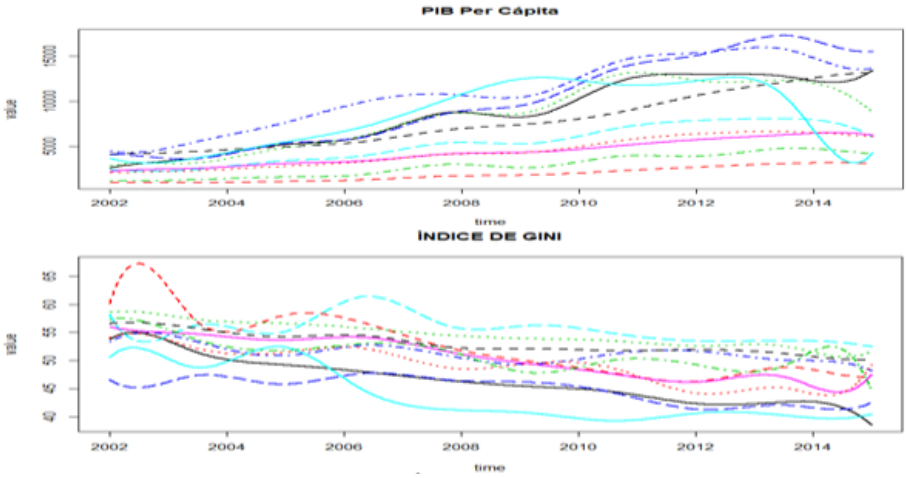


Figura 4: Suavizamiento de las variables explicativas

Como se puede observar en la *Sección* (7.3) las variables explicativas tiene muchos altos y bajos por lo que se suavizan.

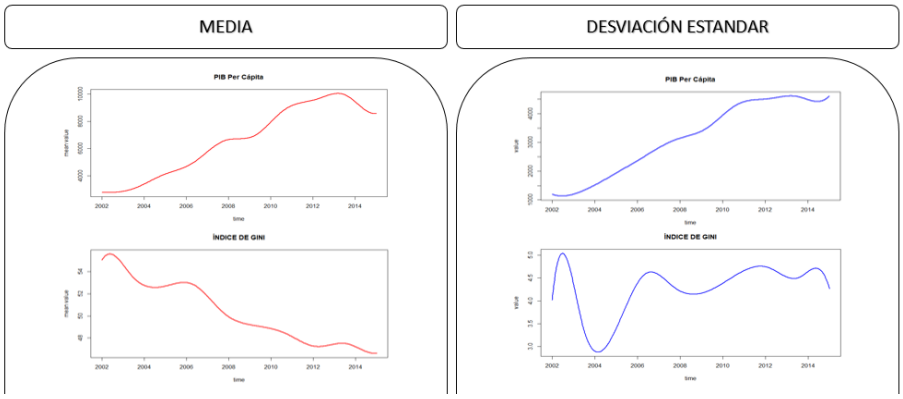


Figura 5: Medias y Varianzas decada variable explicativa a nivel Latinoamérica

Lo que se puede presenciar en la Figura 5 es: (1) En América del Sur la media del PIB Per Cápita a ido ascendiendo a través de los años, pero comienza a decaer a partir del 2013 y porporcionalmente su desviación a ido incrementando. Lo que implica que el nivel de riqueza o bienestar de América del Sur en promedio incrementó hasta el 2012, (2) En promedio el Índice de GINI de América del Sur tiene una tendencia decreciente desde el 2002 hasta el 2015, lo que implica que en general la desigualdad de ingreso a disminuido considerablemente, a pesar de tener una bastante volatividad de la desviación estándar en los años 2002 y 2006 que es más que todo alterada por Bolivia y Venezuela.

7.5. Detección de países atípicos de América del sur

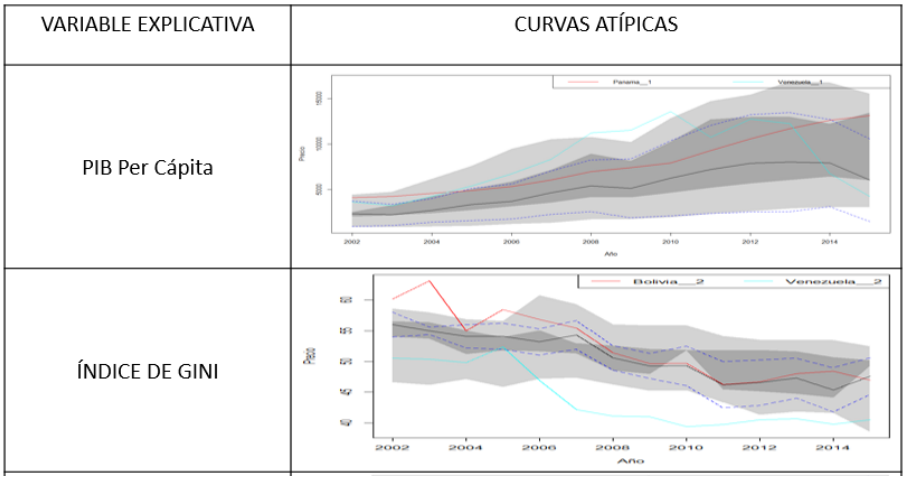


Figura 6: Países atípicos en cada uno de los índices

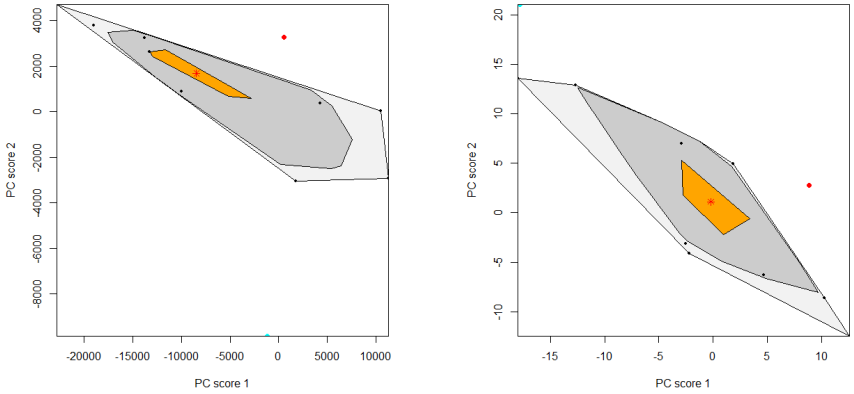


Figura 7: Otra forma de mostrar los países atípicos en cada uno de los índices (PIB Per Cápita y Índice de GINI, respectivamente)

PIB Per Cápita

Hay dos países atípicos en América del Sur, los cuales son:

- **Venezuela:** Este índice es atípico desde el 2008 hasta el 2010, esto se debe a que el presupuesto nacional de 2009 fue calculado estimando el ingreso de 60 dólares por barril de petróleo,[cita requerida] pero a finales de marzo se re-formuló a 40 dólares, para ajustar la caída de los precios del petróleo a nivel global de 2009 y 2010, lo que desencadenó a su vez una crisis energética interna. Luego de la crisis energética, Venezuela sería la única nación petrolera y una de las dos naciones americanas aún en recesión en 2010. El gobierno venezolano culpó a la lenta recuperación económica mundial de alargar la crisis, así como a la reducción de las cuotas de producción petrolera dictadas por la OPEP. En el 2009 país que experimentó un aumento de la deuda pública. Asimismo, cayeron sus reservas internacionales y su Producto Interno

Bruto. Todos estos hechos que dieron lugar a una crisis económica dieron inicio desde el 2008, ya a finales del 2010 la economía comenzó a crecer, por esta razón para el 2011 Venezuela ya no es un dato atípico. http://bit.ly/28QELtf

- **Panamá:** En la Figura 7 se observa que Panamá es atípico en el año 2002 y 2003, luego de la crisis financiera la ganancia sería la más baja desde 2002, cuando el indicador retrocedió; sin embargo sigue siendo de los más altos de América del Sur(*Constanza Morales H., La Tercera. (2016). Panamá desplazaría a Chile como el país de la región con mayor PIB per cápita en 2018.*). http://bit.ly/2nl92dQ
  Desde 2003 hasta 2009 el PIB se duplicó, propiciado por una alta inversión externa e interna, el turismo y la industria logística. Según el Banco Mundial, el FMI y la ONU el país tiene el ingreso por cápital más alto de América Central, el cual es de unos 13 090 dólares; es además el mayor exportador e importador a nivel regional según la CEPAL. El PIB ha crecido de forma sostenida durante más de veinte años seguidos (1989). El país está clasificado en la categoría de «grado de inversión» por parte de las empresas calificadoras de riesgo: Standard and poors, Moody's y Fitch Ratings. http://bit.ly/2aQgyJ5

índice de GINI

Hay dos países atípicos en América del Sur, los cuales son: Bolivia y Venezuela por razones explicadas en la *Sección* (7.3.3).

7.6. Estimación del modelo clásico y Estimación bayesiana del modelo lineal generalizado binomial, por medio del algoritmo seleccionado

Como muchos saben la distribución binomial pasa a ser bernoulli cuando $n = 1$. Por lo tanto en la teoría la explicación se basa en la distribución binomial, con el fin de explicar la parte bayesiana del modelo, por medio del teorema de Bayes.

Modelo lineal generalizado funcional

```
Deviance Residuals:
    Min      1Q  Median      3Q     Max
-1.302  -1.148  -0.940   1.177   1.435

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.134153   0.163859  -0.819    0.413
xc12.bspl4.1 -0.003572   0.008264  -0.432    0.666
xc12.bspl4.2 -0.006620   0.011455  -0.578    0.563
xc12.bspl4.3  0.015411   0.028141   0.548    0.584
xc12.bspl4.4 -0.012050   0.022249  -0.542    0.588
xc12.bspl4.5  0.007613   0.014211   0.536    0.592
xc22.bspl4.1 -2.267498   4.487361  -0.505    0.613
xc22.bspl4.2  2.186938   4.445734   0.492    0.623
xc22.bspl4.3 -1.907968   4.038054  -0.472    0.637
xc22.bspl4.4  0.105727   0.547253   0.193    0.847
xc22.bspl4.5  1.516047   2.915498   0.520    0.603

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.84  on 153  degrees of freedom
Residual deviance: 208.58  on 143  degrees of freedom
AIC: 230.58

Number of Fisher Scoring iterations: 4
```

Modelo lineal generalizado funcional Bayesiano

```
                Mean SD Naive SE Time-series SE
(Intercept) -15.060393  0        0             0
xp1          -0.005424  0        0             0
xp2           0.003177  0        0             0
xp3           0.002300  0        0             0
xp4          -0.003885  0        0             0
xp5           0.010247  0        0             0

      2. Quantiles for each variable:

                    2.5%         25%
(Intercept) -15.060393 -15.060393 -
xp1          -0.005424  -0.005424
xp2           0.003177   0.003177
xp3           0.002300   0.002300
xp4          -0.003885  -0.003885
xp5           0.010247   0.010247

       50%         75%       97.5%
-15.060393 -15.060393 -15.060393
 -0.005424  -0.005424  -0.005424
  0.003177   0.003177   0.003177
  0.002300   0.002300   0.002300
 -0.003885  -0.003885  -0.003885
  0.010247   0.010247   0.010247
```

Figura 8: Estimaciones por el método clásico y Bayesiano

El modelo bayesiano no converge, por lo que se concluye que el mejor modelo es el clásico. Por lo tanto se

explicarán las el modelo escogido en términos de las probabilidades:

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \text{-0,134153} - 0{,}003572 * x_{PIBpc}(t) - 0{,}00662 * x_{PIBpc}(t) + 0{,}015411 * x_{PIBpc}(t) - 0{,}01205 * x_{PIBpc}(t) + 0{,}007613 * x_{PIBpc}(t) - 2{,}267498 * x_{GINI}(t) + 2{,}186938 * x_{GINI}(t) - 1{,}907968 * x_{GINI}(t) + 0{,}105727 * x_{GINI}(t) + 1{,}516047 * x_{GINI}(t)$$

Figura 9: Modelo escogido: Modelo lineal generalizado bernoulli o binomial con n=1

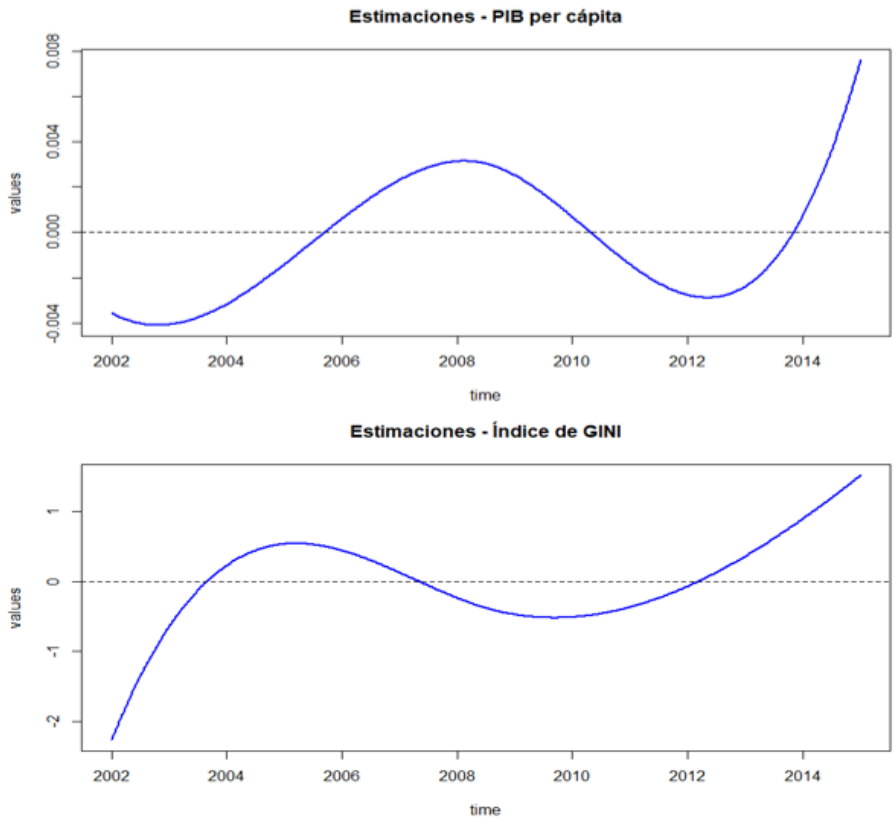


Figura 10: Visualización de las estimaciones

De la *Figura 10* el primer gráfico nos indica que por cada unidad que aumente el PIB Per Cápita teniendo el índice de GINI constante la probabilidad de que un país en Latino América se encuentre en estado de democracia es muy baja desde el año 2002 hasta 2005, un poco alta desde el 2006 hasta el 2010, un poco baja desde el 2011 al 2014 y muy alta para el 2015.

Del segundo gráfico se concluye que por cada unidad que aumente el índice de GINI teniendo el PIB Per Cápita constante la probabilidad de que un país en Latino América se encuentre en estado de democracia es muy baja en el 2002, pero va aumentando, llegando a ser un poco alta la probabilidad desde el 2004 al 2006, un poco baja la probabilidad desde el 2008 hasta el 2012 y va aumentando de probabilidad desde el 2013 al 2015, siendo el 2015 el año con más probabilidad de que en Latino América el estado de un país sea democracia.

7.7. Prueba de independencia del modelo clásico

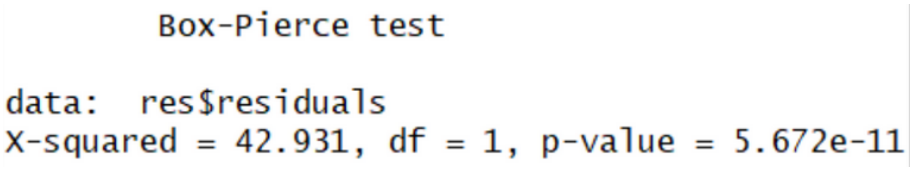


Figura 11: Prueba de independencia

Se rechaza H_0 y se concluye que no independencia entre los residuales.

8. Conclusión

- Gran parte de América del Sur a 2015 tiene un sistema político regido bajo Régimen.
- Venezuela y Bolivia, son países de Latino América que presentan mayor irregularidades económicas, que se ven reflejadas en el índice de GINI y el PIB Per Cápita.
- El modelo bayesiano no converge, por lo que las estimaciones no son efectivas, por lo tanto el mejor modelo es el modelo clásico, ya que sus estimaciones son mejores.
- Hay que tener en cuenta que la variable respuesta (índice de democracia EIU) es un índice mundial, y también que los datos se tomaron como funcionales y al solo analizar los países de solo Latino América la base en si es muy pequeña, ya que ay que tener en cuenta que el análisis funcional es efectivo para grandes volúmenes de datos; lo que implica que el los dos modelos planteados (Bayesiano y Clásico) tendrían mejores resultados si se trabajaran los datos a nivel mundial. Por este motivo como los resultados en la presente tesis no son muy efectivos. Pero se logra identificar que el análisis funcional es bueno con grandes volúmenes de datos y se puede sacar todas las estadísticas que se manejan con una base de datos tomada como multivariada o normal.

Referencias

[1] HORVÁTH, L. y KOKOSZKA, P., (2012) *Inference for Functional Data with Applications*, New York Heidelberg Dordrecht London, (Vol.200) Springer .

[2] RAMSAY, J. O. y SILVERMAN, B.W., (2005) *Functional Data Analysis*, Segunda edición, Springer-Verlag.

[3] HANS-GEORG MÜLLER y ULRICH STADTMÜLLER., (2005) *Generalized functional linear models.*, Project Euclid (mathematics and statistics online). Vol.33 (2) 774-805.

[4] ANDREW GELMAN, JOHN B. CARLIN, HAL S. STERN , DAVID B. DUNSON, AKI VEHTARI y DONALD B. RUBIN, (2014) *Bayesian Data Analysis*,Taylor & Francis Group, Broken Sound Parkway, NW., 3ra Edición.

[5] BANCO MUNDIAL «PIB per cápita (US$ a precios actuales)» https://datos.bancomundial.org/indicador/NY.GDP.PCAP.CD

[6] BANCO MUNDIAL «Índice de Gini».
https://datos.bancomundial.org/indicador/SI.POV.GINI

[7] HANS-GEORG MÜLLER y ULRICH STADTMÜLLER,(2005) «Generalized Functional Linear Models», *The Annals of Statistics*, Institute of Mathematical Statistics, Vol. 33, No. 2, 774-805.

[8] MARÍN, JUAN MIGUEL, (2005)«Tema 1: Introducción a la Estadística Bayesiana». *Departamento de Estadística*, Universidad Carlos III de Madrid.
http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Bayes/tema1bayes.pdf

[9] MAYETRI GUPTA y JOSEPH G. IBRAHIM,(2009) «An Information Matrix Prior for Bayesian Analysis in Generalized Linear Models with High Dimensional Data», Stat Sin. 19(4): 1641-1663.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909687/

[10] FACULTAD DE INGENIERÍA,«ESTADÍSTICA BAYESIANA», Universidad de la República de Uruguay.
https://eva.fing.edu.uy/pluginfile.php/81075/mod_resource/content/1/ESTADISTICA%20BA

[11] CIPRIAN M. CRAINICEANU y A. JEFFREY GOLDSMITH (2010),«Bayesian Functional Data Analysis Using WinBUGS», *Journal of Statistical Software*, Volume 32, Issue 11.
file:///C:/Users/USER/Downloads/v32i11%20(2).pdf