

# UNIVERSITAT POLITÈCNICA DE VALENCIA

Departamento de Matemática Aplicada



Modelización con incertidumbre del incremento  
del uso del comercio electrónico en España  
utilizando técnicas de epidemiología matemática

Ph. D. THESIS

Presentada por: D. Iván Camilo Lombana

Tutores: Dr. Juan Carlos Cortés López  
Dr. Rafael Jacinto Villanueva Micó

Valencia, Junio 2019



Juan Carlos Cortés López y Rafael Jacinto Villanueva Micó, Profesores de la Universitat Politècnica de València,

Certificamos que la presente Tesis, *Modelización con incertidumbre del incremento del uso del comercio electrónico en España utilizando técnicas de epidemiología matemática*, ha sido dirigida bajo nuestra supervisión en el Departamento de Matemática Aplicada de la Universitat Politècnica de València por Iván Camilo Lombana para obtener el título de Doctor en Matemáticas.

Según consta en el informe, en cumplimiento de la legislación vigente, autorizamos la presentación de la tesis doctoral antes mencionada ante la Comisión de Doctorado de la Universitat Politècnica de València, firmando el presente certificado

Valencia, Junio de 2019

Juan Carlos Cortés López

Rafael Jacinto Villanueva Micó



*“Intenta no volverte un hombre de éxito,  
sino volverte un hombre de valor.”*

**(Albert Einstein)**, 1879–1955



*Dedico esta Tesis a*

*Mi Mamá Martha Lombana,*

*Mi Abuela Cecilia Lombana*

FUENTES DE INSPIRACIÓN Y MOTIVACIÓN. . .

*y a mi gran cómplice de vida*

*Yina Enciso*





# Agradecimientos

Agradezco especialmente a *Juan Carlos Cortés y Rafael Jacinto Villanueva*, quienes se encargaron de compartir su conocimiento, fundamentar procesos académicos de investigación, contribuir en este proceso de formación académica y ser ejemplos a seguir como personas y profesionales.

A todas las personas del Instituto Universitario de Matemática Multidisciplinar.

A toda mi familia y amigos.



# Abstract

E-commerce has become one of the accepted purchase systems due to the easy access to much variety of goods, E-commerce provides the possibility to compare them in the terms of costs, features and offers without the need to go to stores and, therefore, save time and money. In addition, the forms of payment are becoming increasingly safe and reliable, which makes the use of electronic commerce a growing practice in its implementation, reason why knowing this growth has to be very useful for discover how purchases can change over the next few years.

In Chapter 1, a mathematical model has been developed to study future short-term trends in electronic commerce in Spain, for this, the total population according to the official data available from the Spanish National Statistics Institute (INE) is divided into 6 age groups from 16 to 74 years old, so that each age group has been divided into two subgroups: the first consists of people who adopt the use of electronic commerce and those who do not. From these subdivisions, a non-linear diffusion model is considered, whose parameters are the innovation coefficients associated with each age group and related to advertising and the imitation coefficients that represent the influence of a group over the others. These parameters are calculated by fitting the available real data of the Spanish National Statistics Institute, as well as an analysis of the sensitivity of the parameters.

In Chapter 2, an age-structured diffusion model is built. It was divided into

two age groups, 15 – 44 and 45 – 74 because the data of INE (2015b), INE (2015a) about electronic commerce users show significant changes between them in the use of this technology. For this model a demographic model was built, then it was scaled due to data are given in percentages. A system of equations in differences was built to describe the demographic evolution in each time instant  $t$  in months for the two age groups. With this system, the parameters of the model that fit the available data are determined to estimate and predict the users of E-Commerce in Spain over the next few years.

In Chapter 3, we retrieve data about the frequent users of electronic commerce during the period 2011-2016 from the Spanish National Institute of Statistics. These data, coming from surveys, have intrinsic uncertainty that we describe using appropriate random variables. Then, we propose a stochastic model to study the dynamics of frequent users of electronic commerce. The goal is to solve the inverse problem that consists of determining the model parameters as suitable parametric random variables, in such a way the model output be capable of capturing the data uncertainty, at the time instants where sample data are available, via adequate probability density functions. To achieve the aforementioned goal, we propose a computational procedure that involves building a nonlinear objective function, based on statistical moment measures, to be minimized using a variation of the particle swarm optimization algorithm (PSO).

# Resumen

El comercio electrónico se ha convertido en uno de los sistemas de compra más aceptados debido a la facilidad para acceder a gran variedad de artículos. Este ofrece la posibilidad de compararlos en términos de costos, características y ofertas sin necesidad de ir a las tiendas y, por tanto, ahorrando tiempo y dinero. Además, las formas de pago se están volviendo cada vez más confiables y seguras, lo que hace del comercio electrónico una práctica creciente en su implementación, razón por la cual conocer este crecimiento ha de ser de gran utilidad para descubrir como las compras pueden cambiar en los próximos años.

En el Capítulo 1, se ha desarrollado un modelo matemático para estudiar futuras tendencias a corto plazo del comercio electrónico en España, para esto, la población total según los datos oficiales disponibles del INE se divide en 6 grupos de edad de 16 a 74 años. A su vez cada grupo de edad se ha dividido en dos subgrupos: el primero consiste en personas que adoptan el uso del comercio electrónico y quienes no lo hacen. A partir de estas subdivisiones se considera un modelo de difusión no lineal cuyos parámetros son los coeficientes de innovación asociados a cada grupo de edad y relacionados con la publicidad, y los coeficientes de imitación que representan la influencia de un grupo hacia los demás. Dichos parámetros se calculan ajustando los datos reales disponibles del Instituto Nacional de Estadística. Además se ha incluido un análisis de estabilidad del modelo y de la sensibilidad de los parámetros.

En el Capítulo 2, se construye un modelo de difusión estructurado por edades. Se divide en dos grupos de edad, 15 – 44 y 45 – 74 porque los datos de (INE, 2015b), (INE, 2015a) sobre los usuarios de comercio electrónico muestran cambios significativos entre dichos grupos de edad en el uso de esta tecnología. Para este modelo se construye un modelo demográfico. Con esta información demográfica se construye un sistema de ecuaciones en diferencias. El modelo fue escalado debido a que los datos estaban dados en porcentajes para describir la evolución demográfica en cada  $t$  en meses para los dos grupos de edad. Con este sistema, los parámetros del modelo que se ajustan a los datos disponibles se determinan para estimar los usuarios del comercio electrónico en España y predecir el uso de esta tecnología para realizar compras en línea.

En el Capítulo 3, recuperamos datos sobre los usuarios frecuentes del comercio electrónico durante el periodo 2011 – 2016 del Instituto Nacional de Estadística. Estos datos, provenientes de encuestas, tienen incertidumbre intrínseca la cual describimos mediante el uso de variables aleatorias apropiadas. Luego, proponemos un modelo estocástico para estudiar la dinámica de los usuarios frecuentes del comercio electrónico. El objetivo es resolver el problema inverso que consiste en determinar los parámetros del modelo como variables aleatorias paramétricas adecuadas, de tal manera que la salida del modelo sea capaz de capturar la incertidumbre de los datos, en los instantes de tiempo en los que se dispone de los datos muestreados, a través de funciones adecuadas de densidad de probabilidad. Para lograr el objetivo mencionado anteriormente, proponemos un procedimiento computacional que implica la creación de una función objetivo no lineal, basada en medidas de los momentos estadísticos, para minimizarla mediante una variación del algoritmo de optimización de enjambre de partículas (PSO).

# Resum

El comerç electrònic s'ha convertit en un dels sistemes de compra més acceptats a causa de la facilitat per a accedir a gran varietat d'articles. Est ofereix la possibilitat de comparar-los en termes de costos, característiques i ofertes sense necessitat d'anar a les botigues i, per tant, estalviant temps i diners. A més, les formes de pagament s'estan tornant cada vegada més de confiança i segures, la qual cosa fa del comerç electrònic una pràctica creixent en la seua implementació, raó per la qual conèixer aquest creixement ha de ser de gran utilitat per a descobrir com les compres poden canviar en els pròxims anys.

En el capítol 1, s'ha desenvolupat un model matemàtic per a estudiar futures tendències a curt termini del comerç electrònic a Espanya, per a això, la població total segons les dades oficials disponibles de l'INE es divideix en 6 grups d'edat de 16 a 74 anys. Al seu torn cada grup d'edat s'ha dividit en dos subgrups: el primer consisteix en persones que adopten l'ús del comerç electrònic i els qui no el fan. A partir d'aquestes subdivisions es considera un model de difusió no lineal els paràmetres de la qual són els coeficients d'innovació associats a cada grup d'edat i relacionats amb la publicitat, i els coeficients d'imitació que representen la influència d'un grup cap als altres. Aquests paràmetres es calculen ajustant les dades reals disponibles de l'Institut Nacional d'Estadística. A més s'ha inclòs una anàlisi d'estabilitat del model i de la sensibilitat dels paràmetres.

En el Capítol 2, es construeix un model de difusió estructurat per edats. Es divideix en dos grups d'edat, 15-44 i 45-74 perquè les dades de INE (2015b), INE (2015a) sobre els usuaris de comerç electrònic mostren canvis significatius entre aquests grups d'edat en l'ús d'aquesta tecnologia. Per a aquest model es construeix un model demogràfic. Amb aquesta informació demogràfica es construeix un sistema d'equacions en diferències. El model va ser escalat pel fet que les dades estaven donades en percentatges per a descriure l'evolució demogràfica en cada  $t$  en mesos per als dos grups d'edat. Amb aquest sistema, els paràmetres del model que s'ajusten a les dades disponibles es determinen per a estimar els usuaris del comerç electrònic a Espanya i predir l'ús d'aquesta tecnologia per a realitzar compres en línia.

En el Capítol 3, recuperem dades sobre els usuaris freqüents del comerç electrònic durant el període 2011-2016 de l'Institut Nacional d'Estadística. Aquestes dades, provinents d'enquestes, tenen incertesa intrínseca la qual descrivim mitjançant l'ús de variables aleatòries apropiades. Després, proposem un model estocàstic per a estudiar la dinàmica dels usuaris freqüents del comerç electrònic. L'objectiu és resoldre el problema invers que consisteix a determinar els paràmetres del model com a variables aleatòries paramètriques adequades, de tal manera que l'eixida del model siga capaç de capturar la incertesa de les dades, en els instants de temps en els quals es disposa de les dades mostrejades, a través de funcions adequades de densitat de probabilitat. Per a aconseguir l'objectiu esmentat anteriorment, proposem un procediment computacional que implica la creació d'una funció objectiu no lineal, basada en mesures dels moments estadístics, per a minimitzar-la mitjançant una variació de l'algorisme d'optimització d'exam de partícules (PSO).



# Índice General

<b>Abstract</b>	<b>XI</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Modelo determinista estructurado por edades para la difusión del comercio electrónico en España</b>	<b>9</b>
2.1. Datos disponibles del comercio electrónico y modelo demográfico estructurado por edades . . . . .	9
2.2. Modelo de difusión . . . . .	12
2.3. Ajuste y predicción . . . . .	15
2.3.1. Ajuste del modelo . . . . .	15
2.3.2. Predicción del uso del comercio electrónico en el corto plazo . . . . .	19
2.4. Análisis de estabilidad y sensibilidad del modelo . . . . .	22
2.4.1. Análisis de incertidumbre de los datos . . . . .	22
2.4.2. Muestreo por Hipercubo Latino (LHS) . . . . .	22
2.4.3. Estabilidad del modelo . . . . .	27
2.5. Conclusiones . . . . .	28
<b>3. Modelo de difusión por edades variando el tamaño de la población para estudiar el uso del comercio electrónico en España</b>	<b>31</b>
3.1. Introducción . . . . .	31
3.2. Construcción del modelo demográfico . . . . .	32

3.3.	Modelo del comercio electrónico . . . . .	33
3.3.1.	Escalado del modelo . . . . .	36
3.3.2.	Ajuste del modelo escalado con datos reales . . . . .	38
3.3.3.	Ajuste probabilístico . . . . .	42
3.3.4.	Procedimiento del ajuste probabilístico . . . . .	44
3.3.5.	Estimación probabilística . . . . .	46
3.3.6.	Resultados . . . . .	48
3.4.	Conclusiones . . . . .	48
<b>4.</b>	<b>Modelo de la dinámica de usuarios del comercio electrónico utilizando técnicas de optimización para problemas inversos con incertidumbre</b>	<b>51</b>
4.1.	Introducción . . . . .	51
4.2.	Datos disponibles y modelo del comercio electrónico . . . . .	54
4.3.	Construcción del modelo para estudiar la dinámica de los usuarios frecuentes del comercio electrónico en España . . . . .	56
4.4.	Procedimiento para resolver el problema utilizando una técnica de optimización inversa . . . . .	59
4.5.	Resultados . . . . .	62
4.6.	Conclusiones . . . . .	66
<b>5.</b>	<b>Conclusiones</b>	<b>69</b>
	<b>Bibliografía</b>	<b>71</b>

# Lista de Tablas

2.1. Datos disponibles correspondiente al porcentaje de población española que han comprado en Internet desde el año 2006 a 2008 (INE, 2008). . . . .	10
2.2. Estimación del parámetro $\mu$ y de las tasas de mortalidad $d_i$ (INE). . . . .	12
2.3. Parámetros de las tasas de crecimiento para el modelo demográfico. . . . .	13
2.4. Estimación de los coeficientes de Innovación por grupos de edad. . . . .	18
2.5. Estimación de los parámetros $C_i$ de <i>contagio</i> . . . . .	18
2.6. Estimación de la susceptibilidad de los parámetros $S_j$ . . . . .	19
2.7. Resultados del análisis de sensibilidad en el año 2012 para los coeficientes de innovación. . . . .	26
2.8. Resultados del análisis de sensibilidad en el año 2012 para coeficientes de imitación. . . . .	27
2.9. Resultados del análisis de estabilidad en el año 2012 para coeficientes de innovación. . . . .	28
2.10. Resultados del análisis de estabilidad en el año 2012 para los coeficientes de imitación. . . . .	28
3.1. Datos del comercio electrónico (CE) de la población española divididos en dos grupos de edad desde 2007 hasta 2015. Obsérvese que la suma de cada fila es el 100% . . . . .	34
3.2. Parámetros que mejor ajustan el modelo con los datos en la Tabla 3.1. . . . .	41

3.3. Predicción para el uso del comercio electrónico desde 2016 a 2019. . . . .	42
3.4. Tamaño de la muestra para cada encuesta considerada. . . . .	43
3.5. Función de probabilidad de cada encuesta. . . . .	44
3.6. Intervalos de confianza (IC) (95 %) para cada muestra. . . . .	47
4.1. Proporción $p_k$ de españoles que usaron el comercio electrónico (CE) al menos una vez en los últimos tres meses durante el periodo 2011 – 2016 y el tamaño de muestra $n_k$ de las encuestas correspondientes (E-commerce, 2018) . . . . .	55
4.2. Distribuciones binomiales escaladas (Bi) para los datos del comercio electrónico de españoles recopilados en la Tabla 4.1 y sus momentos de primer ( $m_k$ ) y segundo orden ( $v_k$ ), $k = 0, 1, \dots, 5$ , definidos en (4.1). . . . .	56

# Índice de Figuras

2.1. Modelo demográfico compartimental para los grupos de edad propuestos. . . . .	11
2.2. Modelo compartimental estructurado por edades para el comercio electrónico en España. . . . .	16
2.3. Representación en 2D de los coeficientes de imitación donde una mayor intensidad en el color indica un mayor valor del coeficiente de imitación. . . . .	20
2.4. Representación en 3D para los coeficientes de imitación. . .	21
2.5. Ajuste y predicción del comercio electrónico en España de 2009 a 2015, $t = 0$ año 2000 y $t = 15$ año 2015. . . . .	23
3.1. Modelo compartimental que representa las poblaciones y las transiciones entre subpoblaciones. . . . .	37
3.2. Representación gráfica de una partícula en el método PSO .	40
3.3. Resultado del ajuste del modelo y predicciones desde 2016 hasta 2019. . . . .	41
3.4. Representación gráfica del ajuste probabilístico. . . . .	45
3.5. Representación gráfica del proceso del ajuste probabilístico.	46
3.6. Representación gráfica de la incertidumbre del modelo y su ajuste. . . . .	49
4.1. Modelo compartimental para la dinámica de la proporción de usuarios frecuentes del comercio electrónico en España. .	57

4.2.	Comparación entre los histogramas de los resultados del modelo $N = 1000$ para $t_1 = \text{Dic } 2012$ , $t_2 = \text{Dic } 2013$ , $t_3 = \text{Dic } 2014$ , $t_4 = \text{Dic } 2015$ , $t_5 = \text{Dic } 2016$ (en azul) y sus histogramas de datos de las distribuciones binomiales escaladas. . .	63
4.3.	Función de densidad de probabilidad del parámetro de transmisión de variable aleatoria $\beta = \text{Be}_p(\beta_1, \beta_2)$ . . . . .	64
4.4.	Histogramas de los resultados del modelo $N = 1000$ en los instantes de tiempo Dic 2017, Dic 2018, Dic 2019 y Dic 2020.	65
4.5.	La media y el intervalo de confianza del 95 % representados por puntos rojos y azules para los datos de Dic. de 2011 a Dic. de 2016. En el eje $X$ denotamos los años y en el eje $Y$ los porcentajes. Las líneas verdes representan el promedio e intervalo de confianza del 95 % para los meses desde Dic. de 2011 hasta Dic. 2020. . . . .	66

# Capítulo 1

## Introducción

Comprender los factores que determinan por qué algunas innovaciones se difunden y otras no, es de interés para las empresas por razones estratégicas, y para los gobiernos para las políticas de desarrollo económico. De particular interés son las respuestas a las siguientes preguntas:

- ¿Cuáles son los factores que afectan la difusión de las innovaciones?
- ¿Son estos factores iguales o diferentes en el mercado primario o secundario?
- ¿Cómo influyen las características del producto, las características ambientales y las políticas públicas en la tasa de difusión en los diferentes mercados?

Para una empresa innovadora que está considerando la introducción de un producto en mercados primarios o secundarios, hay un conjunto de problemas estratégicos en torno a la incertidumbre técnica, la incertidumbre del mercado y la incertidumbre competitiva (Roger and Olaguera, 2003). La incertidumbre técnica se refiere a si el producto se puede fabricar a un coste razonable y dentro de un tiempo razonable. La incertidumbre del mercado se refiere a la viabilidad económica del mercado del producto y la incertidumbre competitiva se relaciona con si otras empresas podrían disminuir el coste del producto.

El segundo conjunto de problemas está relacionado con la aceptación del producto en el mercado. Aquí, la pregunta relevante es si la difusión de la innovación es cuestión de suerte o existen fuerzas identificables que la guían. Parecería evidente que la utilidad de una innovación debería impulsar el proceso de difusión. Es decir, las características de una innovación deben influir en cómo se difundirá. Pero si mantenemos constantes las particularidades de una innovación, existen varias características del entorno que ayudan a determinar la velocidad de difusión. Las características del producto y su utilidad son necesarias pero no son condiciones suficientes para la difusión de una innovación.

Las redes sociales tienen un papel muy importante en el comercio electrónico ya que las empresas las están usando para dar a conocer sus productos y aumentar sus ventas para hacer del uso del comercio electrónico un medio habitual para realizar compras. Estudios de implementación de un modelo de una red social (Lombana et al., 2010) muestran la estructura en la conformación de una red y su importancia en procesos de modelización.

El análisis del proceso de difusión implica un estudio de la innovación, los canales de comunicación, el tiempo y el sistema social. Se puede argumentar que la aceptación o el rechazo de una innovación puede ser más que un mero proceso social asociado con la difusión a través de innovadores, primeros usuarios y demás. Existen algunas estructuras sobre las cuales se debe construir este sistema social para que una idea innovadora, tecnológica o un producto se difunda. Una comprensión de estos entornos estructurales sería bastante valiosa para las empresas al tomar decisiones de introducción de productos.

## **La difusión del comercio electrónico**

La tecnología de difusión es el proceso mediante el cual una innovación se comunica a través de ciertos canales a lo largo del tiempo entre los miembros de un sistema social (Roger and Olaguera, 2003). La formulación



de un modelo matemático confiable de difusión tecnológica debe considerar las características particulares de la tecnología y sus usuarios. A pesar de que varios autores han desarrollado modelos matemáticos de difusión para estudiar algunas tecnologías (Frambach, 1993; Völlink et al., 2006; Zhang and Ntoko, 2002), no consideran aspectos importantes como los diferentes consumidores entre las personas dependiendo de su edad y un mayor impacto de las innovaciones en ciertos grupos de edad. Este aspecto es particularmente importante en relación al uso del comercio electrónico, el cual, requiere tener cierta capacitación en informática, así como habilidades tecnológicas que dependen fuertemente de la edad.

La difusión del comercio electrónico, ha aumentado drásticamente su volumen de negocios en los últimos años debido a sus numerosas ventajas:

- Las tiendas están abiertas las 24 horas del día, los 365 días del año; Internet ofrece la posibilidad de revisar ofertas sin tener que cerrar en cualquier momento.
- Comprar en línea tiene la ventaja de no tener que salir de nuestra casa u oficina para hacer una compra, lo que permite un ahorro del tiempo para otras actividades.
- Permite tener acceso a una gran cantidad y variedad de artículos; antes de poder comprar en línea, nos limitábamos a los artículos ofrecidos por las tiendas locales.
- Ofrece la posibilidad de comprar artículos que no están en nuestra ciudad; además de permitirnos encontrar una variedad de artículos, Internet nos permite productos que pueden estar fuera del área donde vivimos, estado o incluso fuera de nuestro país, lo que aumenta las posibilidades de encontrar artículos de acuerdo con nuestros intereses.
- Permite comparar características y precios inmediatamente; proporciona la capacidad de utilizar servicios de comparación de precios o visitar varios sitios que ofrecen el mismo artículo o similar para com-

parar sus características, precios y costes de envío, y con toda esta información hacer una mejor compra.

En resumen, el comercio electrónico ofrece ahorro de tiempo y evita el desplazamiento para realizar la compra de un artículo, permite la posibilidad de revisión de precios y calidades sin depender de los horarios de una tienda tradicional, permite acceder a una mayor variedad y cantidad de artículos, en muchos casos, con precios más bajos. Además, con la posibilidad de utilizar el servicio de comparación de precios, o visitar varios sitios que ofrecen el mismo artículo o uno similar para comparar sus características, permite al comprador tomar la mejor decisión.

Los modelos de difusión de innovaciones son una herramienta poderosa que tratan de describir, explicar y predecir la reacción colectiva del mercado al introducir una innovación: cómo evoluciona en el tiempo, el número total de usuarios de un nuevo producto o proceso productivo, así como los factores que pueden influir en dicha evolución. Es decir, analizan el proceso a través del cual una innovación se va difundiendo en la sociedad.

El propósito de un modelo de difusión es explicar por medio de modelos matemáticos el incremento o disminución del número de adoptantes a lo largo del proceso, estimar los parámetros de un modelo en su fase inicial, para extrapolar y así realizar estimaciones en el tiempo.

En esta memoria, se ha realizado un estudio sobre la difusión del uso del comercio electrónico en España a partir de datos tomados del INE (Instituto Nacional de Estadística Español). Esta memoria está dividida en tres capítulos. En el Capítulo 1, se plantea un modelo determinista estructurado por edades para el estudio de las futuras tendencias a corto plazo de la difusión del comercio electrónico en España a partir de datos oficiales del INE. En este modelo se asume la población constante, la cual se ha dividido en 6 grupos de edad entre 16 y 74 años en España y se ha estimado el número de consumidores que utilizan Internet para comprar cualquier artículo a lo largo del tiempo. Quienes adoptan la tecnología se pueden clasificar en dos grupos: Los *Innovadores* que deciden usar el comercio electrónico debido al impacto publicitario a través de los medios u otros

factores externos (independientemente de la decisión de los demás) y los *Imitadores*, los consumidores restantes que solo comenzarán a utilizar el comercio electrónico una vez que hayan visto que otras personas lo usan o como resultado de la interacción e influencia de los innovadores sobre los imitadores. Para esto, se ha considerado un modelo de difusión no lineal cuyos parámetros son los coeficientes de innovación, asociados a cada grupo de edad y relacionados con la publicidad, y los coeficientes de imitación, que representan la influencia de un grupo sobre el resto. Estos parámetros se estiman ajustando los datos reales disponibles. En este modelo matemático se ha realizado un análisis de la incertidumbre, ya que los parámetros y condiciones iniciales no se conocen con un grado suficiente de certeza. Esto es debido a la variación natural, el error en las mediciones o simplemente la falta de técnicas para medirlas. El propósito del análisis de incertidumbre es cuantificar el grado de confianza en los datos experimentales existentes y las estimaciones de los parámetros. La técnica utilizada ha sido el muestreo por Hipercubo Latino. También se ha incluido un análisis de la sensibilidad y estabilidad. El primero para estudiar los cambios sobre los resultados (outputs) del modelo ante cambios en los parámetros (inputs) del modelo, y el segundo para cuantificar dichos cambios. El capítulo finaliza con una breve exposición de las principales conclusiones obtenidas a partir del modelo.

A pesar de encontrar buenos ajustes en los parámetros para hacer pronósticos frente al uso del comercio electrónico, se incluyen otros factores en el Capítulo 2, que permitan tener mejores aproximaciones a la realidad con el fenómeno estudiado. Para esto, se cuenta con más datos sobre el uso del comercio electrónico a partir de datos oficiales del INE. La población total se asume variable en el tiempo. En vista de diferencias significativas en ciertos grupos de edad, la población objetivo se divide en 2 subgrupos, 16 – 44 años y 45 – 74 años, donde cada grupo de edad se divide en dos subgrupos, el primero relacionado con personas que adoptan esta tecno-

logía y el segundo que no lo hace. A partir de estos grupos, se construye un sistema no lineal de ecuaciones en diferencias (que describen la evolución de los usuarios del comercio electrónico) cuyos parámetros son los coeficientes de innovación, asociados con cada grupo de edad y relacionados con la publicidad, y los coeficientes de imitación, relacionados con la influencia de un grupo de tecnología adoptante sobre el grupo de no adopción. Estos parámetros se estiman por medio de un ajuste probabilístico. A diferencia del estudio en el capítulo anterior, el proceso de optimización utilizado es el algoritmo de optimización de enjambre de partículas aleatorio (rPSO). Un aspecto clave en el tratamiento del modelo es la estimación de sus parámetros teniendo en cuenta la incertidumbre que existe asociada a los datos del uso del comercio electrónico recogidos en la información del INE. En este capítulo se ha aplicado una técnica, denominada “*Ajuste Probabilístico*”, que permite construir estimaciones puntuales y probabilísticas de los parámetros y, a partir de las mismas, validar el modelo y posteriormente construir predicciones puntuales y probabilísticas del uso del comercio electrónico.

Para el planteamiento y estudio de un modelo matemático de un fenómeno social, la construcción correcta del mismo no es necesariamente suficiente si no se tiene también en cuenta la incertidumbre implícita en los datos de muestra y los parámetros del modelo además de controlar las salidas de estos en la solución. Bajo este enfoque, a partir de datos oficiales del INE sobre usuarios del comercio electrónico en España desde los años 2011 a 2016, el objetivo del Capítulo 3 consiste en incorporar al estudio realizado, las herramientas necesarias para la construcción de un modelo estocástico que permita analizar la dinámica de los usuarios del comercio electrónico en España. Para la construcción de este modelo estocástico, las distribuciones de probabilidad de los parámetros juegan un papel fundamental. Para describir este proceso de forma más realista, es primordial asignar correctamente distribuciones de probabilidad adecuadas a cada uno de los parámetros del modelo. Es por ello que a diferencia de los capítulos ante-

riores, en este último capítulo se aborda el tratamiento de la incertidumbre en el modelo de la dinámica del comercio electrónico. Ahora el objetivo es determinar las distribuciones de probabilidad más adecuadas para describir la incertidumbre del fenómeno. Con este enfoque se pretende calcular predicciones a través de la determinación de la distribución de probabilidad de la salida (output) del modelo en cada instante temporal, en lugar de restringirnos a la obtención de predicciones puntuales y probabilísticas mediante la media e intervalos de confianza, respectivamente. Más específicamente, el objetivo es resolver los problemas inversos a partir de la estimación de los parámetros del modelo como variables aleatorias paramétricas adecuadas, de tal forma que la salida del modelo sea capaz de capturar la incertidumbre de los datos a través de distribuciones de probabilidad adecuadas desde el momento en que se dispone de los datos muestreados. Para ello, se construye un procedimiento computacional que permita construir una función objetivo no lineal basado en medidas de los momentos estadísticos, la cual debe ser minimizada utilizando una variación del algoritmo de optimización implementado en el Capítulo 2 denominado enjambre de partículas o Particle Swarm. El capítulo concluye con una análisis de los resultados y las principales conclusiones que se derivan del estudio de modelización realizado.

*Introducción*

---

## Capítulo 2

# Modelo determinista estructurado por edades para la difusión del comercio electrónico en España

### 2.1. Datos disponibles del comercio electrónico y modelo demográfico estructurado por edades

Se ha considerado para el estudio datos oficiales del Instituto Nacional de Estadística (INE) en España sobre el porcentaje de españoles que compraron en Internet de 2006 a 2008, (INE, 2008). Estos datos, los únicos disponibles en el momento de realizar el estudio, se clasificaron por grupos de edad como se muestra en la Tabla 2.1.

Las características sociales de los individuos y sus edades influyen en el proceso de difusión de una innovación; en este caso, los factores de mayor aceptación en el uso de esta tecnología (comercio electrónico) dependen de la edad, la familiaridad con su uso y las habilidades. Se propone un modelo de población por edades basado en dos estados hacia la innovación: quienes

Tabla 2.1: Datos disponibles correspondiente al porcentaje de población española que han comprado en Internet desde el año 2006 a 2008 (INE, 2008).

Grupos de edad	Porcentaje en 2006	Porcentaje en 2007	Porcentaje en 2008
16 - 24 años	28.7	28.7	32.2
25 - 34 años	31.6	33.4	37.9
35 - 44 años	23.2	24.3	26.5
45 - 54 años	14.6	18	20
55 - 64 años	7.2	7.8	8.6
65 - 74 años	1.3	2	2

adoptan la tecnología (comercio electrónico) y quienes no la adoptan. Para quienes adoptan la tecnología, distinguimos dos posibilidades: La primera corresponde a los *Innovadores*, quienes adoptan el comercio electrónico influenciados por las campañas publicitarias o el comportamiento individual del consumidor. La segunda posibilidad son los *Imitadores*, quienes incorporan esta tecnología solo por interacción social o relación confiable entre otros que usan el comercio electrónico, luego influenciados por ellos, los imitadores deciden usar la tecnología.

Teniendo en cuenta los datos oficiales del INE, es necesario construir un modelo demográfico consistente y estructurado por edades para integrar esta información en el modelo de difusión. Este modelo demográfico estructurado por edades considera no solo el intervalo de edad de acuerdo con la Tabla 2.1, sino también las tasas de crecimiento correspondientes y las tasas de mortalidad implícitas. De esta forma, se definen los grupos de edad de la siguiente forma:

- Grupo 1 ( $G_1$ ) : Población de personas entre 16 y 24 años,
- Grupo 2 ( $G_2$ ) : Población de personas entre 25 y 34 años,
- Grupo 3 ( $G_3$ ) : Población de personas entre 35 y 44 años,
- Grupo 4 ( $G_4$ ) : Población de personas entre 45 y 54 años,



- Grupo 5 ( $G_5$ ) : Población de personas entre 55 y 64 años,
- Grupo 6 ( $G_6$ ) : Población de personas entre 65 y 74 años.

El modelo demográfico por edades que se ha considerado está dado por el siguiente sistema de ecuaciones diferenciales ordinarias y representado por la Figura 2.1. :

$$\begin{aligned}
 G_1'(t) &= \mu - c_1 G_1(t) - d_1 G_1(t), \\
 G_j'(t) &= c_{j-1} G_{j-1}(t) - c_j G_j(t) - d_j G_j(t), \quad 2 \leq j \leq 5, \\
 G_6'(t) &= c_5 G_5(t) - d_6 G_6(t).
 \end{aligned}
 \tag{2.1}$$

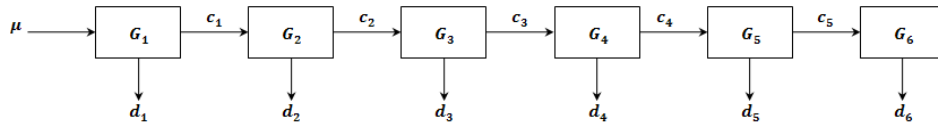


Figura 2.1: Modelo demográfico compartimental para los grupos de edad propuestos.

De los datos del INE, la población promedio entre 16 – 74 años es 33.955.126. En este caso, la *tasa de nacimientos* ( $\mu$ ) corresponde a los jóvenes que ingresan al sistema, es decir, las personas de 16 años de edad y las tasas de mortalidad por grupo ( $d_i$ ,  $i = 1, \dots, 6$ ) de edad están dadas por la Tabla 2.2.

El coeficiente  $d_6$  se ha ajustado teniendo en cuenta que en el último grupo de edad, a partir además de la muerte de la población, hay una proporción de personas que cambiaron a otro grupo de edad (75 años o más) que no se ha considerado en la población objetivo. Esta tasa se calcula por (2.1). Asumiremos que la población total así como la de cada grupo de edad permanecen constantes en el tiempo (Hethcote, 2000). Esta suposición ha sido utilizado por otros autores y simplifica el estudio, y es razonable en países como España. Como consecuencia se puede suponer que:

Tabla 2.2: Estimación del parámetro  $\mu$  y de las tasas de mortalidad  $d_i$  (INE).

Parámetros	Estimación
$\mu$	0.013131184
$d_1$	0.000528376
$d_2$	0.000625189
$d_3$	0.001259272
$d_4$	0.002774752
$d_5$	0.006132741
$d_6$	0.102301753

$$G'_1(t) = G'_2(t) = G'_3(t) = G'_4(t) = G'_5(t) = G'_6(t) = 0. \quad (2.2)$$

Una vez considerados los coeficientes  $\mu$  y  $d_i$ , a partir de los de la Tabla 2.2, se denotan los parámetros  $c_k$ ,  $k = 1, \dots, 5$  (no se considera la tasa de crecimiento  $c_6 = 0$ , porque el grupo 6 es la subpoblación final para la población objetivo). Para determinar las tasas de crecimiento de cada grupo de edad, se resuelve el sistema (2.1) a partir de (2.2), luego se despeja  $c_k$ ,  $k = 1, \dots, 5$  de cada ecuación y se obtiene:

$$c_1 = \frac{\mu}{G_1} - d_1, c_j = c_{j-1} \frac{G_{j-1}}{G_j} - d_j, 2 \leq j \leq 5. \quad (2.3)$$

Al reemplazar las estimaciones de los parámetros  $\mu$  y las tasas de mortalidad  $d_i$  en (2.3) se obtienen las tasas de crecimiento para cada grupo de edad dados en la Tabla 2.3.

## 2.2. Modelo de difusión

En este apartado, se propone un modelo matemático de difusión para estudiar el comercio electrónico en España. En primer lugar, se presentan dos subpoblaciones las cuales clasifican a las personas con edades comprendidas entre 16 – 74 años para cada uno de los 6 subintervalos:

Tabla 2.3: Parámetros de las tasas de crecimiento para el modelo demográfico.

Tasas de crecimiento por grupo de edad	Estimación
$c_1$	0.096208836
$c_2$	0.057813091
$c_3$	0.058894496
$c_4$	0.069227322
$c_5$	0.080844739
$c_6$	0

- $N_i(t)$ ,  $i = 1, \dots, 6$ , denota la tasa de población del grupo  $i$ -ésimo que no adopta la tecnología, en el tiempo  $t$ .
- $Y_i(t)$ ,  $i = 1, \dots, 6$ , denota la tasa de población del grupo  $i$ -ésimo que adopta la tecnología, en el tiempo  $t$ .

La difusión de la tecnología estará representada por el paso de un individuo en la población  $N_i(t)$  a  $Y_i(t)$  a través de los coeficientes de innovación o imitación descritos por:

- $P_i$ ,  $i = 1, \dots, 6$ , es el coeficiente de innovación para el grupo de edad  $i$ -ésimo.
- $C_i$ ,  $i = 1, \dots, 6$ , mide la intensidad con la cual el grupo de edad  $i$ -ésimo influencia (o *contagia*) el uso de la tecnología (comercio electrónico) a otros grupos de edad.
- $S_j$ ,  $j = 1, \dots, 6$ , mide la susceptibilidad de ser convencido a adoptar esta tecnología por los individuos que pertenecen al  $j$ -ésimo grupo de edad influenciados por otras personas que usan el comercio electrónico.
- $Q_{ij} = C_i S_j$ ,  $i, j = 1, \dots, 6$ , representa la tasa de transmisión del uso de la tecnología por encuentros entre individuos del grupo  $i$ -ésimo que

ya han adoptado la tecnología (comercio electrónico) y las personas del  $j$ -ésimo grupo que aún no la han adoptado. Para cada grupo el coeficiente de imitación  $Q_{ij}$  se factoriza en dos factores,  $C_i$  y  $S_j$ . De esta forma reducimos sustancialmente el número de parámetros del modelo.

Además, se tienen en cuenta las siguientes consideraciones:

- Se supone una población homogénea, es decir, cada individuo puede adoptar el comercio electrónico por contacto con cualquier persona que usa el comercio electrónico, con la misma probabilidad (Hethcote, 2000).
- Se considera que las personas de 16 años acceden a la subpoblación  $Y_1(t)$  con una tasa dada por  $\mu \frac{Y_1(t)}{N_1(t)+Y_1(t)}$  si han adoptado el uso de la tecnología (comercio electrónico) o  $N_1(t)$  con una tasa  $\mu \frac{N_1(t)}{N_1(t)+Y_1(t)}$  si no lo han hecho aún. Es decir, la entrada de nuevos miembros al sistema se distribuye en el mismo % de usuarios y no usuarios del primer grupo de edad.
- Hay dos maneras en que un individuo que pertenece a  $N_i(t)$  transite a  $Y_i(t)$ : La primera forma es por medio de un coeficiente de innovación (representa la tasa de adquisición de un nuevo producto independientemente de las decisiones del resto de las personas), que está relacionado con la influencia publicitaria, el comportamiento individual del consumidor, etc., el cual está modelizado por  $P_i N_i(t)$ ,  $i = 1, \dots, 6$ . La segunda forma es por medio de un comportamiento de imitación, que se modeliza a través de un coeficiente de imitación. Éste representa la transmisión del uso del comercio electrónico por encuentros entre individuos de los grupos  $Y_i(t)$  que ya han adoptado la tecnología (comercio electrónico) e individuos de  $N_i(t)$  que aún no lo han adoptado. Está dado por un término no lineal modelizado por  $N_i(t) \sum_{j=1}^6 Q_{ij} Y_j(t)$ , es decir, teniendo en cuenta la posibilidad de contagio con usuarios del comercio electrónico de cualquier edad.

- Una persona que ha adoptado el comercio electrónico, siempre permanecerá en el estado de individuos que ya han comprado por Internet y puede influenciar a personas que aún no lo han hecho.

Bajo los supuestos anteriores y teniendo en cuenta que  $c_6 = 0$ , se construye el modelo matemático estructurado por edades del comercio electrónico en España utilizando técnicas de epidemiología matemática basado en el sistema no lineal de ecuaciones diferenciales ordinarias dado por (2.4) – (2.9). La Figura 2.2 muestra una representación compartimental de este sistema.

$$N_1'(t) = \mu \frac{N_1(t)}{N_1(t) + Y_1(t)} + (-c_1 - d_1 - P_1)N_1(t) - N_1(t) \sum_{j=1}^6 Q_{1j}Y_j, \quad (2.4)$$

$$Y_1'(t) = \mu \frac{Y_1(t)}{N_1(t) + Y_1(t)} + P_1N_1(t) + (-c_1 - d_1)Y_1(t) + N_1(t) \sum_{j=1}^6 Q_{1j}Y_j, \quad (2.5)$$

$$N_i'(t) = c_{i-1}N_{i-1}(t) + (-c_i - d_i - P_i)N_i(t) - N_i(t) \sum_{j=1}^6 Q_{ij}Y_j, \quad 2 \leq i \leq 5, \quad (2.6)$$

$$Y_i'(t) = c_{i-1}Y_{i-1}(t) + P_iN_i(t) + (-c_i - d_i)Y_i(t) + N_i(t) \sum_{j=1}^6 Q_{ij}Y_j, \quad 2 \leq i \leq 5, \quad (2.7)$$

$$N_6'(t) = c_5N_5(t) + (-d_6 - P_6)N_6(t) - N_6(t) \sum_{j=1}^6 Q_{6j}Y_j, \quad (2.8)$$

$$Y_6'(t) = c_5Y_5(t) + P_6N_6(t) - d_6Y_6(t) + N_6(t) \sum_{j=1}^6 Q_{6j}Y_j. \quad (2.9)$$

## 2.3. Ajuste y predicción

### 2.3.1. Ajuste del modelo

En la Sección 2.1, se han obtenido los parámetros  $\mu$ ,  $d_i$  y  $c_k$ , sin embargo, resta calcular los coeficientes  $P_i$ ,  $C_i$  y  $S_j$ ,  $i, j = 1, \dots, 6$ . La estimación de

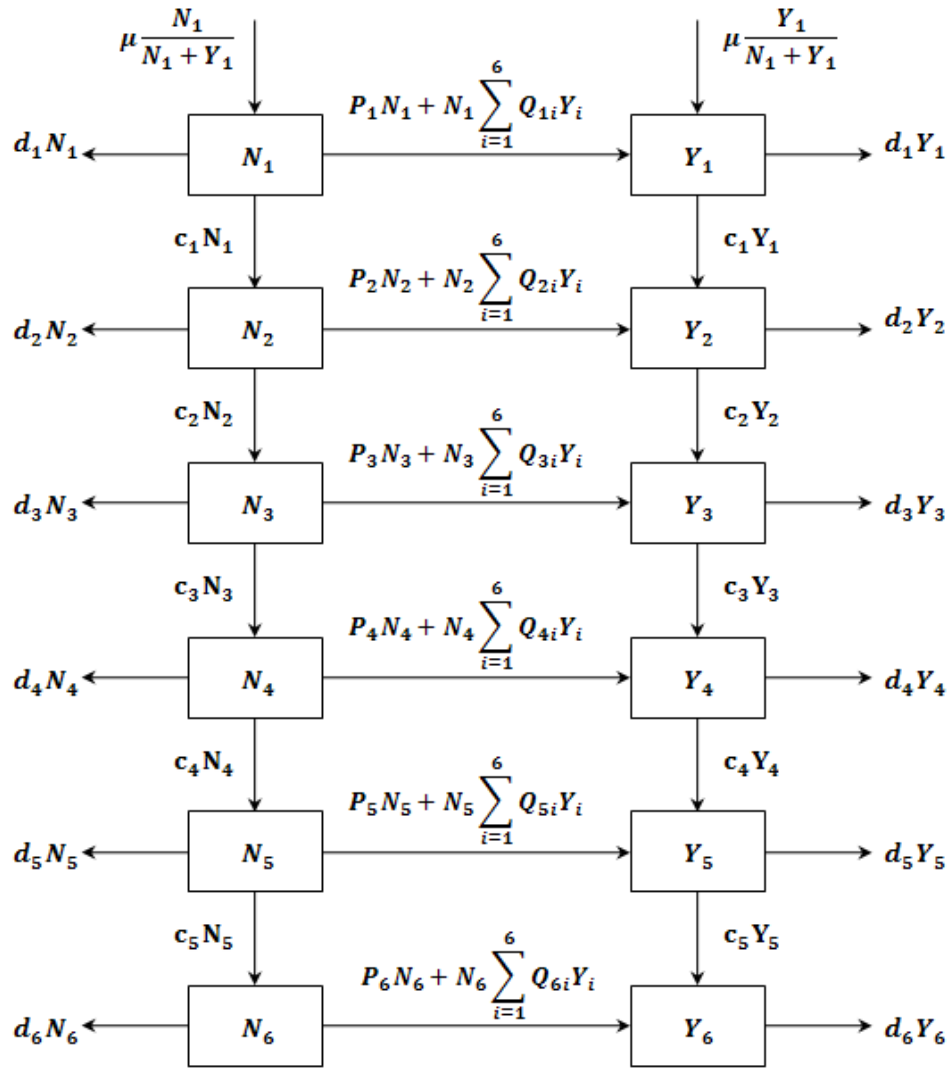


Figura 2.2: Modelo compartimental estructurado por edades para el comercio electrónico en España.

estos dieciocho parámetros debe realizarse de tal forma que la solución del correspondiente sistema de ecuaciones diferenciales no lineal (2.4) – (2.9) proporcione los valores más cercanos a los datos disponibles. La bondad de la aproximación se mide mediante la media cuadrática.

Para este objetivo, en primer lugar se supone que  $P_i$ ,  $C_i$ ,  $S_j$  son no negativos y se genera una muestra aleatoria de estos dieciocho parámetros, a continuación se resuelve el sistema resultante (2.4) – (2.9) por medio de *Mathematica*<sup>®</sup> (Mathematica, 2008) y el comando `NDSolve [ ]` que resuelve numéricamente el sistema (2.4) – (2.9) y se calcula el error cuadrático medio entre el output del modelo en 2006, 2007 y 2008 y los datos correspondientes a los años 2006, 2007 y 2008. Se han realizado 10.000 iteraciones para este procedimiento utilizando el algoritmo de Nelder-Mead (Nelder and Mead (1964)) que nos permite determinar el mínimo global de la función de error

$$E : \mathbb{R}^{18} \longrightarrow \mathbb{R}^+ \cup \{0\},$$

que representa el menor error cuadrático medio, cuyo valor obtenido fue 0,0035. Con respecto a las condiciones iniciales, se ha tomado el año 2000 (que corresponde a  $t = 0$ ) como el inicio del uso del comercio electrónico en España (Mas (2004)). Además se han asignado los valores  $Y_2(2000) = Y_3(2000) = 0.01$  porque asumimos, en general, que las personas pertenecientes a los grupos  $G_2$  y  $G_3$  estaban más activas con las compras en Internet.

Los resultados del coeficiente de innovación  $P_i$  proporcionan información crucial para adoptar estrategias publicitarias adecuadas (véase la Tabla 2.4). Se puede inferir que los grupos  $G_1 - G_3$  son los más influenciados por la publicidad, mientras que las futuras campañas para incentivar el uso del comercio electrónico debería enfocarse al resto de los grupos que tienen coeficientes de innovación más bajos (véase la Tabla 2.4).

La estimación para los coeficientes  $C_i$  y  $S_j$ , que determinan los coeficientes de imitación  $Q_{ij} = C_i S_j$  se han representado en la Figura 2.3. Una mayor intensidad en la Figura 2.3 indica un mayor valor del coeficiente de

Tabla 2.4: Estimación de los coeficientes de Innovación por grupos de edad.

Coefficientes de Innovación por grupos de edad	Estimación
$P_1$	0.0465057
$P_2$	0.0263457
$P_3$	0.0256832
$P_4$	$7.06868 \times 10^{-6}$
$P_5$	0.00488839
$P_6$	0.00045393

Tabla 2.5: Estimación de los parámetros  $C_i$  de contagio.

Parámetros de Contagio $C_i$	Estimación
$C_1$	0.038669
$C_2$	0.678257
$C_3$	0.209166
$C_4$	0.494765
$C_5$	0.053162
$C_6$	0.002184

imitación. La representación en 3D viene dada por la Figura 2.4. Teniendo en cuenta la Tabla 2.4, la Tabla 2.5 y la Figura 2.3, se puede asegurar:

- Los resultados obtenidos para los grupos  $G_5$  y  $G_6$  son bajos ya que no están familiarizados con el uso de esta tecnología y para  $G_1$  porque, en general, no tienen un poder adquisitivo significativo que pueda influir en los demás.
- Las personas que pertenecen a los grupos  $G_2$  y  $G_3$  no están influenciadas por las otras, ya que tienen su opinión formada sobre el comercio



Tabla 2.6: Estimación de la susceptibilidad de los parámetros  $S_j$ .

Susceptibilidad de los Parámetros $S_j$	Estimación
$S_1$	1.211120
$S_2$	0.146283
$S_3$	0.154951
$S_4$	0.918915
$S_5$	0.973203
$S_6$	0.634599

electrónico principalmente a través de las campañas de publicidad y compras. Por otro lado, los grupos  $G_2 - G_4$  ejercen una influencia importante sobre el grupo  $G_1$  debido a que son personas adultas que tienen más experiencia con el uso de esta tecnología e incluso en muchos casos tienen una relación familiar, por ejemplo, son familiares de personas del grupo  $G_1$ .

- Finalmente, los grupos  $G_2$  y  $G_3$  también influyen en los grupos  $G_4$  y  $G_5$  porque las personas entre 45 y 64 años confían no solo en ellos sino también en las personas de 25 a 34 años que tienen potencial de compra al igual que buenas habilidades tecnológicas.

### 2.3.2. Predicción del uso del comercio electrónico en el corto plazo

En las dos apartados anteriores se han determinado los valores de los parámetros que proporcionan la mejor solución (en media cuadrática) del sistema no lineal (2.4) – (2.9) con respecto a los datos oficiales de INE (2006 – 2008) sobre el comercio electrónico en España. Esta solución está representada en la Figura 2.5. A pesar de tener pocos datos, el ajuste obtenido es muy bueno. Además, este enfoque permite pronosticar las tendencias

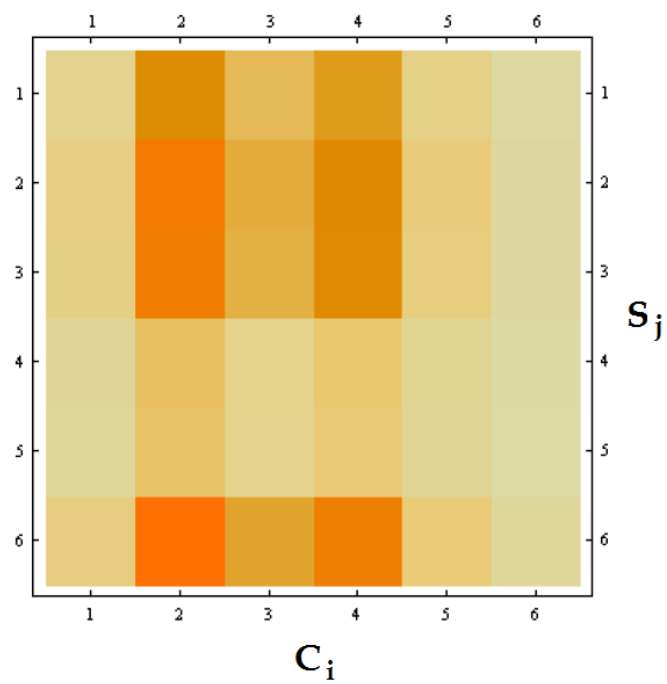


Figura 2.3: Representación en 2D de los coeficientes de imitación donde una mayor intensidad en el color indica un mayor valor del coeficiente de imitación.

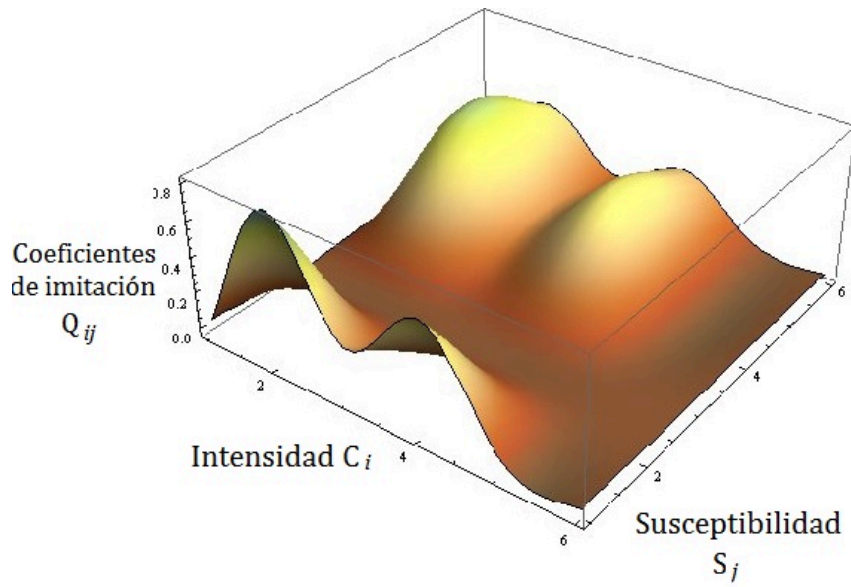


Figura 2.4: Representación en 3D para los coeficientes de imitación.

futuras de esta tecnología en el corto plazo. De la Figura 2.5, inferimos un comportamiento creciente de adopción del comercio electrónico en los próximos años, especialmente para personas de los grupos  $G_1 - G_4$ .

## 2.4. Análisis de estabilidad y sensibilidad del modelo

### 2.4.1. Análisis de incertidumbre de los datos

En la formulación de modelos matemáticos basados en un sistema de ecuaciones diferenciales, como el propuesto para describir la dinámica del comercio electrónico en España, los elementos de entrada o inputs son las condiciones iniciales y los parámetros del modelo. En la práctica estos datos no son conocidos de forma precisa, y suelen contener errores procedentes del muestreo o de las mediciones requeridas para fijar estos datos o por la propia complejidad del fenómeno objeto de estudio. El propósito del análisis de incertidumbre es cuantificar el grado de confianza en los datos experimentales existentes y las estimaciones de los parámetros. Un enfoque muy utilizado para realizar un análisis de la incertidumbre, se denomina Hipercubo Latino, y está basado en una técnica de muestreo tipo Monte Carlo.

### 2.4.2. Muestreo por Hipercubo Latino (LHS)

El muestreo por Hipercubo Latino pertenece a la clase Monte Carlo de métodos de muestreo. LHS permite una estimación no sesgada de simulaciones del modelo, con la ventaja que requiere menos muestras que el muestreo aleatorio simple para lograr la misma precisión. LHS es una técnica también llamada *muestreo estratificado sin reemplazo*.

El análisis de estabilidad y sensibilidad para cuantificar el rango de variabilidad en la salida del modelo (estabilidad) y para identificar los factores claves que dan lugar a los resultados del modelo (sensibilidad) se han vuelto esenciales para determinar la solidez y fiabilidad del modelo.

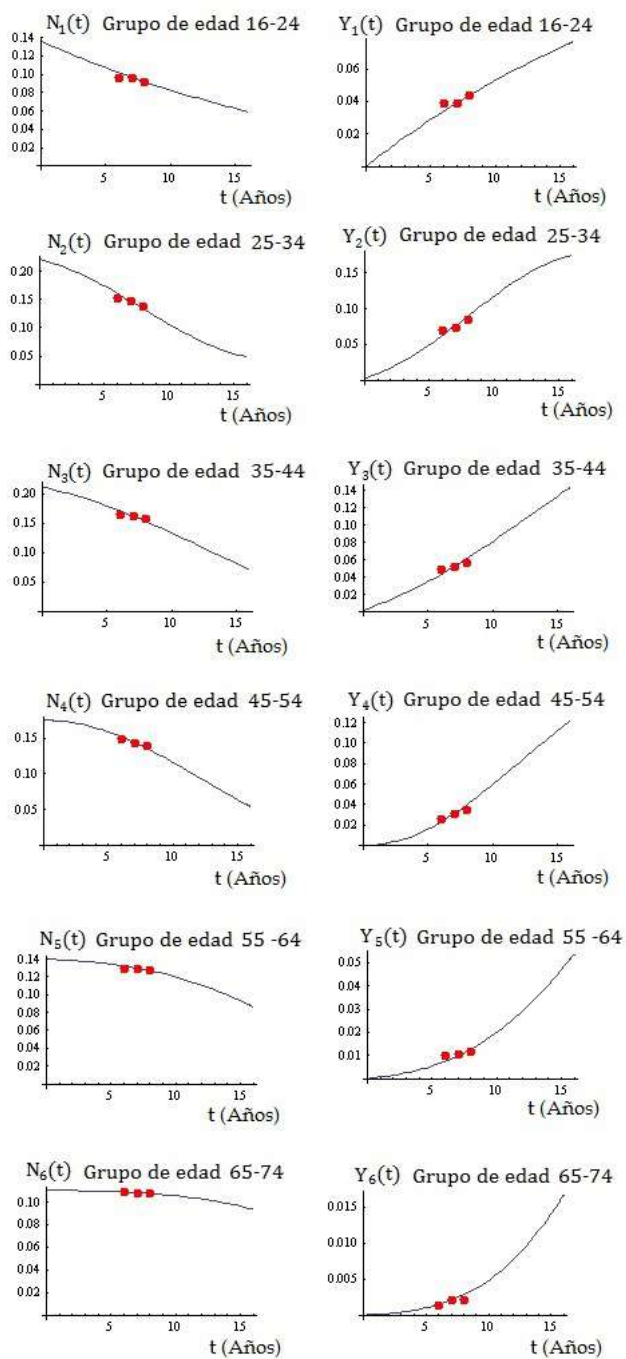


Figura 2.5: Ajuste y predicción del comercio electrónico en España de 2009 a 2015,  $t = 0$  año 2000 y  $t = 15$  año 2015.

Los análisis de estabilidad y sensibilidad se utilizan para determinar el grado de incertidumbre en los resultados del modelo que se debe a la incertidumbre en los parámetros de entrada (parámetros de transmisión). Cada parámetro de entrada para un modelo se puede definir para que tenga asociada una función de densidad de probabilidad apropiada. En este caso, se ha utilizado una distribución uniforme centrada en estimadores de parámetros determinísticos por la ausencia de datos para informar sobre la distribución para un parámetro dado (Marino et al., 2008; Hoare et al., 2008). Por tanto, el modelo se puede simular muestreando un único valor de cada distribución de parámetros. Se deben tomar muchas muestras y deben ejecutarse muchas simulaciones, produciendo valores de salida. El muestreo del LHS es un tipo de muestreo de Monte Carlo estratificado, es un método sofisticado y eficiente para lograr un muestreo equitativo de todos los parámetros de entrada simultánea (Marino et al., 2008; Hoare et al., 2008; Blower and H., 1994).

El análisis de sensibilidad utiliza los parámetros de transmisión generados por LHS como entradas del modelo, de modo que a partir de cada muestra de los parámetros y condiciones iniciales del modelo generada vía LHS se obtiene un resultado (output) distinto del modelo. Mediante un análisis de la sensibilidad se puede determinar los valores de los parámetros y condiciones iniciales cuya variación afecta más a la respuesta del modelo.

Aunque existen diferentes versiones del LHS, en nuestro estudio hemos aplicado la siguiente versión (Olsson et al., 2003):

1. Se denota  $N$  como el número de muestras y  $K$  el número de entradas aleatorias (condiciones iniciales y/o parámetros del modelo).
2.  $P$  es una matriz  $N \times K$  en la que cada una de las columnas  $K$  es una permutación aleatoria de  $1, \dots, N$  y  $R$  es una matriz  $N \times K$  de números aleatorios independientes de la distribución  $(0, 1)$  uniforme.
3. Se calcula  $S$  como  $S = (P - R)/N$ .
4. Cada elemento de  $S = (s_{ij})$  se mapea según su distribución de proba-

bilidad marginal como  $x_{ij} = F_j^{-1}(s_{ij})$ ,  $F^{-1}$  representa el inverso de la función de distribución de probabilidad acumulada para la entrada  $j$ .

En el análisis de estabilidad, una versión de LHS que consigue una reducción de la correlación espúrea. Aunque la distribución marginal de cada entrada se representa de manera eficiente, existe el riesgo que aparezca una correlación espúrea. Sin embargo, se ha demostrado Owen (1994) que dicha correlación espúrea puede reducirse mediante modificaciones en la matriz  $P$  que aparece en el procedimiento anterior. Para calcular la modificación en la definición de matriz  $P$ , seguimos el siguiente procedimiento:

1.  $P$  es una matriz  $N \times K$  en la que cada una de las columnas  $K$  es una permutación aleatoria de  $1, \dots, N$ .
2. Se dividen las entradas de  $P$  por  $N + 1$  (número de muestras o realizaciones más uno) y se mapean con una distribución gaussiana de media cero y desviación estándar uno. Se denota esta nueva matriz por  $Y$ .
3. Se estima la matriz de covarianza de  $Y$  y se calcula la factorización de Cholesky de  $Y$ . Entonces,  $cov(Y) = LL^T$ .
4. Se calcula una nueva matriz  $Y^\dagger = Y(L^{-1})^T$ . Los rangos\* de los elementos de las columnas de  $Y^\dagger$  se convierten en los elementos en las columnas de la nueva matriz  $P$ .

En resumen, se ha utilizado el método LHS para muestrear el espacio de parámetros de entrada, luego se han utilizado estas muestras para ejecutar múltiples evaluaciones del modelo. Estas muestras junto con las salidas generadas por el modelo se han implementado para realizar un análisis de sensibilidad por coeficientes de correlación de rango parcial (PRCC). Se

---

\*Si una columna de  $Y^\dagger$  es 0,2,5,4,1,2,0,01, la columna correspondiente de  $P$  es 2,4,3,1, porque 0,2 es el segundo en relación a la columna de  $Y^\dagger$ , y así sucesivamente.

pueden realizar pruebas de significación para evaluar si una correlación es significativa, es decir, si un PRCC es significativamente diferente de cero. El nivel de confianza (DeGroot and Schervish, 2016) utilizado es del 95 %. Por lo tanto, con este método se puede cuantificar qué parámetro puede desestabilizar el modelo (análisis de estabilidad) y qué parámetros son los más importantes para la variabilidad en el resultado del modelo (análisis de sensibilidad) (Marino et al. (2008); Hoare et al. (2008); Blower and H. (1994)).

Es conveniente tomar un número suficientemente alto de muestras  $N$  (Marino et al. (2008)), en nuestro caso, se toman  $N = 2500$ , se aplica el algoritmo del procedimiento anterior, se calculan las salidas de la matriz y se desarrolla un análisis de sensibilidad por PRCC para estudiar la sensibilidad o la estabilidad del modelo.

Para validar los procesos realizados, se construye el LHS y se elabora un análisis de sensibilidad por medio de PRCC.

Con los resultados obtenidos, se calcula la correlación media, la varianza de las correlaciones obtenidas, los percentiles 5 y 95, y el porcentaje de veces que las correlaciones han sido significativas (véase Tablas 2.7 y 2.8).

Tabla 2.7: Resultados del análisis de sensibilidad en el año 2012 para los coeficientes de innovación.

Parámetro $P_i$	Promedio de sensibilidad	% de veces de sensibilidad significativa
$P_3$	0.00103	6.2
$P_6$	0.0003	5.2
$P_4$	0.00046	4.8
$P_5$	0.00017	4.6
$P_2$	-0.00118	3.3
$P_1$	-0.00004	0.2

Los resultados en las Tablas 2.7 y 2.8 muestran que los parámetros más



sensibles son, por este orden,  $P_3$ ,  $C_3$ ,  $C_1$  y  $C_6$ , ya que tienen el porcentaje de veces de sensibilidad significativa más altas frente a los otros parámetros. Este resultado puede interpretarse del siguiente modo: si se realizan variaciones sobre estos parámetros, este cambio afectará en mayor medida a los resultados (outputs) del modelo que si hay variaciones del mismo tamaño sobre el resto de parámetros del modelo.

Tabla 2.8: Resultados del análisis de sensibilidad en el año 2012 para coeficientes de imitación.

Parámetro	Sensibilidad Promedio	Porcentaje	Parámetro	Sensibilidad Promedio	Porcentaje
$C_1$	-0.00026	5.6	$S_1$	0.00008	5.6
$C_2$	0.00088	5.1	$S_2$	-0.00004	3.5
$C_3$	0.00012	6.3	$S_3$	0.00037	5.2
$C_4$	0.00042	5.3	$S_4$	0.00014	4.0
$C_5$	0.00054	3.6	$S_5$	-0.00034	5.4
$C_6$	0.00175	5.6	$S_6$	-0.00064	3.5

### 2.4.3. Estabilidad del modelo

Para estudiar la estabilidad del modelo (2.4) – (2.9), se aplica el procedimiento descrito en el Apartado 2.4.2 y se construye la matriz de entrada con el número de muestras  $N = 2500$  y la cantidad de parámetros  $K = 18$ .

Para cada muestra (filas de matriz de entrada), se calcula la solución del modelo, es decir, el porcentaje de personas para cada grupo de edad que acepta o no la nueva tecnología para  $t = 2012$  y  $t = 2015$ . De este modo, se construye la matriz de salida.

A partir de las matrices de entrada como de salida, se calcula la matriz PRCC y la matriz de valores de significatividad. El procedimiento descrito se ha repetido 1000 veces usando *Mathematica*.

Tabla 2.9: Resultados del análisis de estabilidad en el año 2012 para coeficientes de innovación.

Parámetro $P_i$	Sensibilidad Promedio	Porcentaje
$P_1$	0.00011	0.1
$P_2$	0.00042	5.1
$P_3$	-0.00039	5.3
$P_4$	-0.00073	4.6
$P_5$	0.00106	5.7
$P_6$	-0.00078	5.1

Tabla 2.10: Resultados del análisis de estabilidad en el año 2012 para los coeficientes de imitación.

Parámetro $C_i$	Sensibilidad Promedio	Porcentaje	Parámetro $S_j$	Sensibilidad Promedio	Porcentaje
$C_1$	0.00024	4.5	$S_1$	0.00037	5.0
$C_2$	-0.00049	5.0	$S_2$	-0.00108	3.2
$C_3$	0.00138	6.3	$S_3$	-0.00023	3.9
$C_4$	-0.00002	5.3	$S_4$	0.00018	5.1
$C_5$	-0.00011	3.7	$S_5$	-0.00127	3.3
$C_6$	-0.0007	5.1	$S_6$	-0.00033	4.1

## 2.5. Conclusiones

En este capítulo se ha propuesto un modelo para describir, en el corto-medio plazo, la dinámica de la evolución del uso del comercio electrónico en España a partir de datos del INE (Instituto Nacional de Estadística).

El modelo se ha basado en un sistema de ecuaciones diferenciales no lineales inspirado en un enfoque epidemiológico para explicar la dinámica de esta nueva tecnología. Para realizar el estudio, y atendiendo a la información disponible a través del INE, la población se ha dividido en 6 grupos de edad. El modelo resultante depende de ciertos parámetros que se han determinado teniendo en cuenta los valores reales disponibles a través del INE. Los resultados obtenidos han sido muy buenos, lo que nos ha permitido realizar predicciones que muestran una clara tendencia del uso del comercio electrónico en los próximos años. En el estudio se ha incluido un análisis de sensibilidad para determinar los parámetros del modelo que, ante pequeños cambios, provocarán una mayor fluctuación en la respuesta o output del modelo. De este modo se ha obtenido información que puede ser clave para conocer cuáles son las subpoblaciones que, ante cambios que pueden estar provocados por alteraciones en su comportamiento de consumo o al impacto recibido por campañas de marketing, más podría afectar a las predicciones realizadas sobre el uso del comercio electrónico.



## Capítulo 3

# Modelo de difusión por edades variando el tamaño de la población para estudiar el uso del comercio electrónico en España

### 3.1. Introducción

En el capítulo anterior hemos considerado muchos grupos de edad y el comportamiento en muchos de ellos es similar. Además, queremos avanzar en el tratamiento de la incertidumbre, lo que implica un aumento en el coste computacional.

En este capítulo, se analizará el fenómeno de la difusión del comercio electrónico con el objetivo de modelar su dinámica suponiendo que el uso del comercio electrónico se debe a una transmisión (contagio) entre pares, es decir, porque los individuos en el entorno convencen a otros de usar el comercio electrónico.

De esta forma, se han considerado dos grupos de edad, 15 – 44 y 45 – 74 porque los datos de INE (2015b), INE (2015a) sobre los usuarios del comercio electrónico arrojan diferencias significativas entre ellos. De esta forma, utilizando datos demográficos, se establece un modelo demográfico en el que varía el tamaño de la población y a partir de él se construye un modelo dinámico que permite determinar la evolución de los usuarios del comercio electrónico a lo largo del tiempo. Para este tipo de modelos determinísticos discretos, las herramientas matemáticas más adecuadas para analizarlos son las ecuaciones y los sistemas de ecuaciones en diferencias. Una ecuación en diferencias es un sistema dinámico que describe la evolución de alguna variable a lo largo del tiempo.

### **3.2. Construcción del modelo demográfico**

El objetivo de esta sección es doble, en primer lugar se construye el modelo demográfico (3.1), y en segundo lugar se propone un modelo matemático para describir la dinámica del comercio electrónico en España, (Sección 3.3). Este último modelo será escalado, ya que los datos sobre usuarios del comercio electrónico para los grupos de edad propuestos en el estudio están dados en porcentajes.

Teniendo en cuenta los datos oficiales de INE (INE, 2015a), se requiere un modelo demográfico estructurado por edades con un tamaño de población variable. Este modelo demográfico se integra en el modelo de difusión del comercio electrónico. Este modelo estructurado por edades considera dos grupos de edad diferentes debido a las diferencias significativas entre ambos grupos con respecto a su comportamiento con el uso del comercio electrónico. Estos grupos son:

- Grupo 1 ( $G_1(t)$ ): Población entre 15 y 44 años en el tiempo  $t$  (en meses).
- Grupo 2 ( $G_2(t)$ ): Población entre 45 y 74 años en el tiempo  $t$  (en meses).

El siguiente sistema de ecuaciones en diferencias (Hethcote (2000)) describe la evolución demográfica para cada  $t$  en meses para los dos grupos de edades diferentes:

$$\begin{aligned}G_1(t+1) &= \mu P_T - c_1 G_1(t) - d_1 G_1(t), \\G_2(t+1) &= c_1 G_1(t) - d_2 G_2(t),\end{aligned}\tag{3.1}$$

donde  $P_T = G_1 + G_2$ ,  $\mu$  es el número de nacimientos mensual (teniendo en cuenta que en España casi nadie muere entre 0 y 14 años),  $c_1$  es la tasa de crecimiento mensual de  $G_1$  a  $G_2$ ,  $d_1$  es la tasa de mortalidad mensual en el primer grupo,  $d_2$  es la tasa de personas entre 45 – 74 años de edad que salen del modelo, por fallecimiento o por cumplir más de 74 años. Es importante destacar que el estudio de los datos disponibles proviene de INE (2015a) y para garantizar que los parámetros representen la realidad tanto como sea posible, el valor de  $\mu$  tiene que estar en el intervalo  $[7,5125 \cdot 10^{-4}, 9,0475 \cdot 10^{-4}]$ ,  $d_1$  en el intervalo  $[3,73 \cdot 10^{-5}, 5,97 \cdot 10^{-5}]$  y  $d_2$  en el intervalo  $[3,3849 \cdot 10^{-3}, 4,12 \cdot 10^{-3}]$ .

Se considera que el tamaño de la población en el modelo demográfico varía con el tiempo. Las técnicas para manejar este tipo de modelos en versiones continuas se han presentado en (Martcheva and Castillo-Chavez, 2003) y (Mena-Lorca and Hethcote, 1992), de forma que hemos adaptado estas técnicas a modelos discretos.

### 3.3. Modelo del comercio electrónico

En este apartado, se propone un modelo discreto que describe la dinámica de la compra por comercio electrónico. Además también se recogen los datos del INE (INE, 2008) relacionados con el uso del comercio electrónico en España en los dos grupos de edad diferentes. Usando el modelo descrito junto con los datos disponibles en INE (2015a) y en INE (2015b), es posible determinar los parámetros del modelo que mejor se ajustan a los datos. Este modelo se construye a partir del modelo demográfico (3.1).

Uso del comercio electrónico	Grupo 1 (15 – 44 años)		Grupo 2 (45 – 74 años)	
	No usuario CE	Usuario CE	No usuario CE	Usuario CE
$t_1 = \text{Dic } 2007 (j = 1)$	0.4255	0.1416	0.3824	0.0400
$t_2 = \text{Dic } 2008 (j = 2)$	0.3955	0.1790	0.3822	0.0431
$t_3 = \text{Dic } 2009 (j = 3)$	0.3652	0.2039	0.3755	0.0551
$t_4 = \text{Dic } 2010 (j = 4)$	0.3425	0.2158	0.3781	0.0607
$t_5 = \text{Dic } 2011 (j = 5)$	0.3242	0.2284	0.3730	0.0742
$t_6 = \text{Dic } 2012 (j = 6)$	0.2891	0.2546	0.3716	0.0845
$t_7 = \text{Dic } 2013 (j = 7)$	0.2668	0.2661	0.3718	0.0951
$t_8 = \text{Dic } 2014 (j = 8)$	0.2258	0.2958	0.3568	0.1214
$t_9 = \text{Dic } 2015 (j = 9)$	0.1891	0.3230	0.3459	0.1417

Tabla 3.1: Datos del comercio electrónico (CE) de la población española divididos en dos grupos de edad desde 2007 hasta 2015. Obsérvese que la suma de cada fila es el 100 %

En la Tabla 3.1, se pueden encontrar datos recuperados del INE (INE, 2015b) sobre los usuarios y no usuarios del comercio electrónico, por grupo de edad, desde 2007 hasta 2015 en España. En esa encuesta, los no usuarios se han definido como aquellos que no han usado el comercio electrónico, al menos, en los últimos 3 meses. Los datos disponibles nos llevan a definir los siguientes subgrupos:

- $N_1(t)$ ,  $N_2(t)$ , corresponden al número de personas que no han usado el comercio electrónico en los últimos 3 meses en el instante  $t$ , para grupos de edad  $G_1(t)$  y  $G_2(t)$ , respectivamente.
- $Y_1(t)$ ,  $Y_2(t)$  corresponden al número de personas que han usado el comercio electrónico en los últimos 3 meses, en el instante  $t$ , para grupos de edad  $G_1(t)$  y  $G_2(t)$ , respectivamente.

Vamos a considerar que una persona utiliza el comercio electrónico porque:

1. Él/ella lo decide por convicción propia (innovador).



2. Él/ella es influenciado por otras personas que lo usan (imitador).

La difusión de la tecnología estará representada por la transición de un individuo de la población  $N_i(t)$  a  $Y_i(t)$  ( $i = 1, 2$ ) a través de los coeficientes de innovación o imitación descritos por:

- $p_1, p_2$  son los coeficientes de innovación para los grupos de edad  $G_1$  y  $G_2$ , respectivamente.
- $\alpha_1, \alpha_2$  son los coeficientes a transmisión de hábitos en  $N_1$  por individuos en  $Y_1$  y  $Y_2$ , respectivamente.
- $\alpha_3, \alpha_4$  son los coeficientes a transmisión de hábitos en  $N_2$  por individuos en  $Y_1$  y  $Y_2$ , respectivamente.

Si un individuo que ya ha comprado a través de Internet deja de comprar en un tiempo de 3 meses, cambia a  $N_1(t)$  o  $N_2(t)$  dependiendo de su edad. Este comportamiento se describirá en el modelo por los coeficientes  $\gamma_1, \gamma_2$ . Específicamente  $\gamma_1$  y  $\gamma_2$  están relacionados con personas de edades comprendidas entre 15 – 44 y 45 – 74, respectivamente.

Además, se consideran los siguientes supuestos:

- El paso del tiempo se define por meses.
- Cada persona puede contactar con cualquier otra, es decir, suponemos que hay una mezcla de población homogénea (Hethcote, 2000).
- Ya que el número de muertes de personas menores de 14 años es no significativa, se supone que esas personas nunca han comprado por Internet (ya que a pesar que los jóvenes con 15 años muestran buenas habilidades con el uso de internet, son un grupo de edad que no tienen vida crediticia que les permita hacer compras online, además no hay estudios oficiales por parte del INE que establezca el uso del comercio electrónico por parte de ellos). Con esta suposición podemos establecer que las personas con 15 años ingresan a  $N_1(t)$  con una tasa  $\mu$ , donde  $\mu$  es la tasa de nacimiento mensual.

- La población total  $P_T(t) := N_1(t) + Y_1(t) + N_2(t) + Y_2(t)$  es variable a lo largo del tiempo.

Bajo los supuestos anteriormente descritos, se construye el modelo de difusión matemática estructurado por edades que describe el uso del comercio electrónico en España con un tamaño de población variable utilizando técnicas de epidemiología matemática con incertidumbre basado en el siguiente sistema no lineal de ecuaciones en diferencias:

$$N_1(t+1) = (1-d_1)N_1(t) + \gamma_1 Y_1(t) - c_1 N_1(t) - N_1(t) \frac{\alpha_1 Y_1(t) + \alpha_2 Y_2(t)}{P_T(t)} - p_1 N_1(t) + \mu P_T(t), \quad (3.2)$$

$$Y_1(t+1) = (1-d_1)Y_1(t) - \gamma_1 Y_1(t) - c_1 Y_1(t) + N_1(t) \frac{\alpha_1 Y_1(t) + \alpha_2 Y_2(t)}{P_T(t)} + p_1 N_1(t), \quad (3.3)$$

$$N_2(t+1) = (1-d_2)N_2(t) + \gamma_2 Y_2(t) + c_1 N_1(t) - N_2(t) \frac{\alpha_3 Y_1(t) + \alpha_4 Y_2(t)}{P_T(t)} - p_2 N_2(t), \quad (3.4)$$

$$Y_2(t+1) = (1-d_2)Y_2(t) - \gamma_2 Y_2(t) + c_1 Y_1(t) + N_2(t) \frac{\alpha_3 Y_1(t) + \alpha_4 Y_2(t)}{P_T(t)} + p_2 N_2(t). \quad (3.5)$$

Una representación gráfica del modelo anterior se puede ver en la Figura 3.1.

### 3.3.1. Escalado del modelo

Teniendo en cuenta que la Tabla 3.1 los datos están dados en porcentajes, es necesario escalar el modelo. En primer lugar, para establecer una relación entre  $P_T(t)$  y  $P_T(t+1)$ , se suman las relaciones (3.2) a (3.5) obteniendo:

$$P_T(t+1) = P_T(t) + \mu P_T(t) - d_1(N_1(t) + Y_1(t)) - d_2(N_2(t) + Y_2(t)). \quad (3.6)$$

Si denotamos por

$$n_1(t) := \frac{N_1(t)}{P_T(t)}, \quad y_1(t) := \frac{Y_1(t)}{P_T(t)}, \quad n_2(t) := \frac{N_2(t)}{P_T(t)}, \quad y_2(t) := \frac{Y_2(t)}{P_T(t)}, \quad (3.7)$$

y se divide (3.2) por  $P_T(t+1)$ , se obtiene la siguiente expresión:

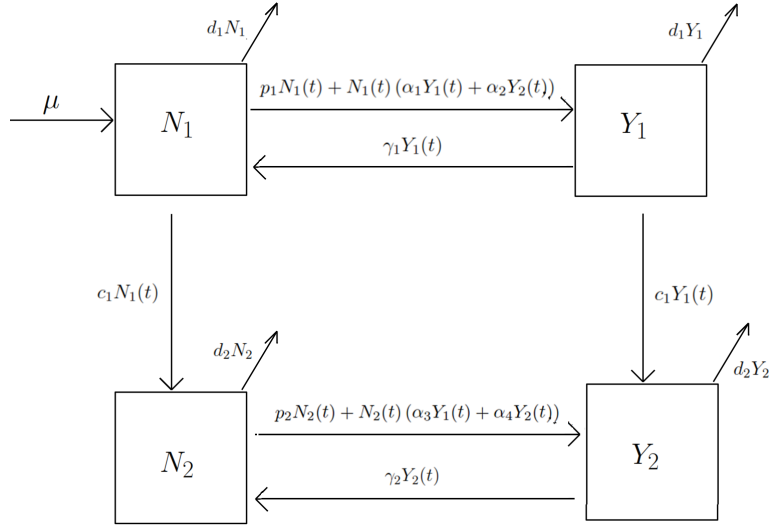


Figura 3.1: Modelo compartimental que representa las poblaciones y las transiciones entre subpoblaciones.

$$\frac{N_1(t+1)}{P_T(t+1)} = \frac{\mu P_T(t) + (1 - d_1)N_1(t) + \gamma_1 Y_1(t) - c_1 N_1(t) - N_1(t) \frac{\alpha_1 Y_1(t) + \alpha_2 Y_2(t)}{P_T(t)} - p_1 N_1(t)}{P_T(t) + \mu P_T(t) - d_1(N_1(t) + Y_1(t)) - d_2(N_2(t) + Y_2(t))}. \quad (3.8)$$

Dividiendo numerador y denominador de la parte derecha de (3.8) por  $P_T(t)$ , se obtiene:

$$n_1(t+1) = \frac{n_1(t) + \mu - d_1 n_1(t) + \gamma_1 y_1(t) - n_1(t)(\alpha_1 y_1(t) + \alpha_2 y_2(t)) - c_1 n_1(t) - p_1 n_1(t)}{1 + \mu - d_1(n_1(t) + y_1(t)) - d_2(n_2(t) + y_2(t))}. \quad (3.9)$$

Aplicando la misma técnica en (3.3) a (3.5), se obtiene el sistema escalado de ecuaciones en diferencias dado por:

$$n_1(t+1) = \frac{n_1(t) + \mu - d_1 n_1(t) + \gamma_1 y_1(t) - n_1(t) (\alpha_1 y_1(t) + \alpha_2 y_2(t)) - c_1 n_1(t) - p_1 n_1(t)}{1 + \mu - d_1 (n_1(t) + y_1(t)) - d_2 (n_2(t) + y_2(t))}, \quad (3.10)$$

$$y_1(t+1) = \frac{y_1(t) - d_1 y_1(t) - \gamma_1 y_1(t) + n_1(t) (\alpha_1 y_1(t) + \alpha_2 y_2(t)) - c_1 y_1(t) + p_1 n_1(t)}{1 + \mu - d_1 (n_1(t) + y_1(t)) - d_2 (n_2(t) + y_2(t))}, \quad (3.11)$$

$$n_2(t+1) = \frac{n_2(t) - d_2 n_2(t) + \gamma_2 y_2(t) - n_2(t) (\alpha_3 y_1(t) + \alpha_4 y_2(t)) + c_1 n_1(t) - p_2 n_2(t)}{1 + \mu - d_1 (n_1(t) + y_1(t)) - d_2 (n_2(t) + y_2(t))}, \quad (3.12)$$

$$y_2(t+1) = \frac{y_2(t) - d_2 y_2(t) - \gamma_2 y_2(t) + n_2(t) (\alpha_3 y_1(t) + \alpha_4 y_2(t)) + c_1 y_1(t) - p_2 n_2(t)}{1 + \mu - d_1 (n_1(t) + y_1(t)) - d_2 (n_2(t) + y_2(t))}. \quad (3.13)$$

### 3.3.2. Ajuste del modelo escalado con datos reales

Una vez que se ha escalado el modelo, el objetivo consiste en encontrar los parámetros

$$X = (\mu, d_1, c_1, d_2, p_1, \alpha_1, \alpha_2, \gamma_1, p_2, \alpha_3, \alpha_4, \gamma_2)$$

que hagan que el modelo se ajuste lo mejor posible a los datos en la Tabla 3.1.

Por un lado, se denota por  $n_1^*(t_j)$ ,  $y_1^*(t_j)$ ,  $n_2^*(t_j)$  y  $y_2^*(t_j)$ ,  $j = 1, \dots, 9$ , la primera, la segunda, la tercera y la cuarta columna en la Tabla 3.1, respectivamente. Por otro lado, denotemos por  $n_1(t_j, X)$ ,  $y_1(t_j, X)$ ,  $n_2(t_j, X)$  y  $y_2(t_j, X)$ , el modelo evaluado utilizando el conjunto de parámetros  $X$  en cada instante de tiempo  $t_j$ , ( $j = 1, \dots, 9$ ). Las condiciones iniciales del modelo vienen dadas por la primera fila de la Tabla 3.1, así que  $n_1(t_1, X) = n_1^*(t_1)$ ,  $y_1(t_1, X) = y_1^*(t_1)$ ,  $n_2(t_1, X) = n_2^*(t_1)$  y  $y_2(t_1, X) = y_2^*(t_1)$ .

La función de error está definida por:

$$\begin{aligned} \text{Error}(X) &= \sum_{j=2}^9 |n_1^*(t_j) - n_1(t_j, X)| + \sum_{j=2}^9 |y_1^*(t_j) - y_1(t_j, X)| \\ &+ \sum_{j=2}^9 |n_2^*(t_j) - n_2(t_j, X)| + \sum_{j=2}^9 |y_2^*(t_j) - y_2(t_j, X)|. \end{aligned} \quad (3.14)$$

A continuación se ha utilizado el algoritmo de optimización de enjambre de partículas aleatorio (rPSO) (Jacob and Khemka, 2004) para minimizar

(3.14), obteniendo el conjunto de parámetros,  $X^*$ , que mejor ajustan el modelo a los datos dados. Los resultados se recogen en la Tabla 3.2.

El método de optimización por enjambre de partículas (conocido como PSO, por sus siglas en inglés, *Particle Swarm Optimization*), es un método heurístico que fue desarrollado para solucionar problemas de optimización continuos, pero recientemente se ha investigado la forma de implementarlo en problemas discretos a través de métodos basados en operadores de intercambio, transformaciones del espacio, matrices difusas y algoritmos híbridos. Fue descrito por primera vez por (Eberhart and Kennedy, 1995). En un principio fueron concebidos para elaborar modelos de conductas sociales, como el movimiento descrito por los organismos vivos en una bandada de aves o un banco de peces. Posteriormente el algoritmo se simplificó y se comprobó que era adecuado para problemas de optimización. PSO permite optimizar un problema a partir de una población de soluciones candidatas, denotadas como partículas que se considera están en movimiento por todo el espacio de búsqueda según reglas matemáticas que tienen en cuenta la posición y la velocidad de las partículas. El movimiento de cada partícula se ve influenciado por su mejor posición local hallada hasta el momento, así como por las mejores posiciones globales encontradas por otras partículas a medida que recorren el espacio de búsqueda. El fundamento teórico de esto, es hacer que la nube de partículas converja rápidamente hacia las mejores soluciones.

El funcionamiento del método es el siguiente: Cada partícula (individuo) tiene una posición,  $\vec{p}$  en el espacio de búsqueda y una velocidad  $\vec{v}$  con la que se mueve a través del espacio. Estas partículas tienen una cantidad de inercia, que los mantienen en la misma dirección en la que se movían, así como una aceleración, que depende de:

- Cada partícula es atraída hacia la mejor localización que ella ha tenido (mejor personal).
- Cada partícula es atraída hacia la mejor localización que ha sido

encontrada por el conjunto de partículas en el espacio global (mejor global).

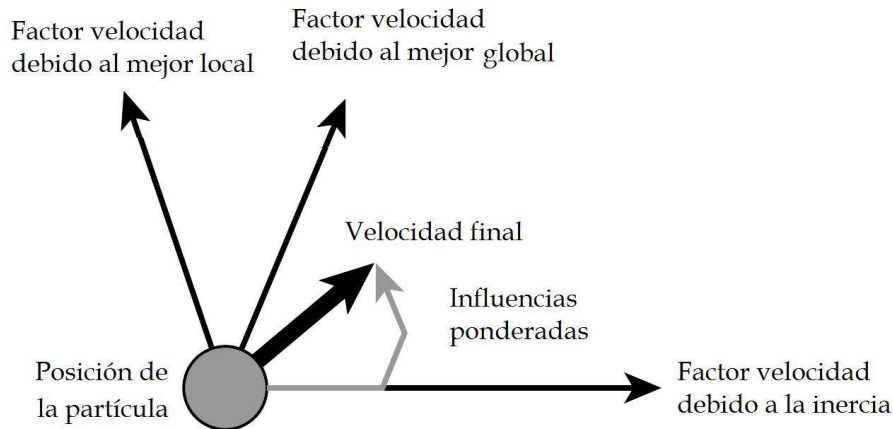


Figura 3.2: Representación gráfica de una partícula en el método PSO .

La fuerza con que las partículas son empujadas en cada una de estas direcciones depende de dos parámetros que pueden ajustarse (atracción-al-mejor-personal y atracción-al-mejor-global), de forma que a medida que las partículas se alejan de estas localizaciones mejores, la fuerza de atracción es mayor. También se suele incluir un factor aleatorio que influye en cómo las partículas son empujadas hacia estas localizaciones.

Este método de optimización es especialmente adecuado cuando el coste de la función de ajuste es muy elevado, pues requiere pocas evaluaciones para encontrar valores cercanos al óptimo, y no requiere que la función objetivo sea diferenciable.

En nuestro caso, tenemos una función en búsqueda de su mínimo global. Nombramos la función objetivo de tal forma que tenga como variable un vector. Cada variable delimitada en un enrejado que tenga una cota mínima y otra máxima. Se construye una función auxiliar que evite que las partículas no sobrepasen los límites demarcados. En caso que una partícula sobrepase los límites marcados, se sustituye por un valor aleatorio por los

límites demarcados inicialmente. Se halla el mejor ajuste encontrado en las partículas que determina el mínimo global.

$\mu$	$7.2650 \cdot 10^{-4}$	$\gamma_1$	0.001553	$\gamma_2$	0.901928
$d_1$	$5.0333 \cdot 10^{-5}$	$\alpha_1$	0.030568	$\alpha_3$	0.598622
$d_2$	$6.2792 \cdot 10^{-5}$	$\alpha_2$	0.011360	$\alpha_4$	0.286072
$c_1$	$1.8602 \cdot 10^{-3}$	$p_1$	0.000106	$p_2$	0.022511

Tabla 3.2: Parámetros que mejor ajustan el modelo con los datos en la Tabla 3.1.

En la Figura 3.3 se ha representado el modelo evaluado según los parámetros en la Tabla 3.2. Los puntos rojos representan los datos en la Tabla 3.1. Además, las predicciones para los próximos 4 años que se muestran en la Tabla 3.3 se han trazado en la misma figura.

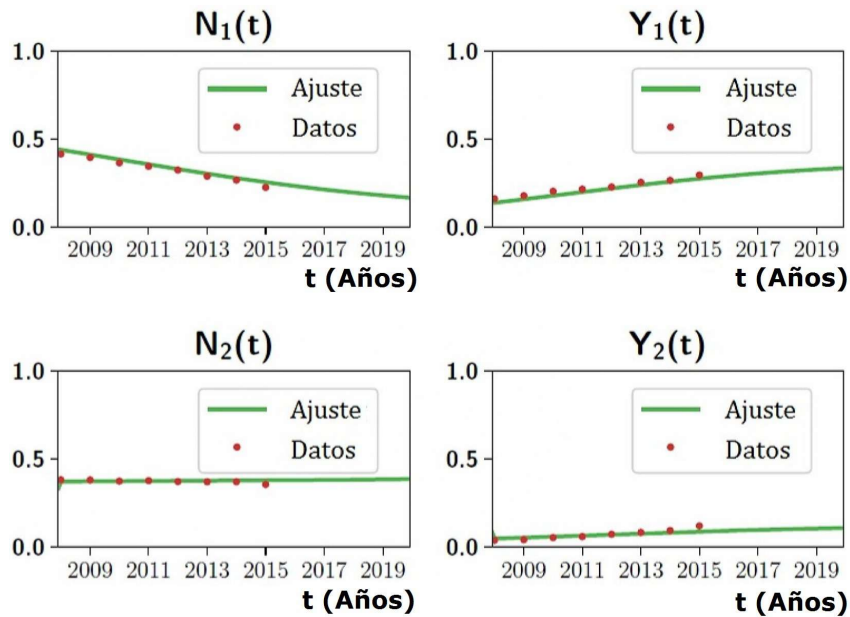


Figura 3.3: Resultado del ajuste del modelo y predicciones desde 2016 hasta 2019.

Usuarios comercio electrónico	Grupo 1 (15 – 44 años)		Grupo 2 (45 – 74 años)	
	No	Si	No	Si
Dic 2016 ( $j = 1$ )	0.21548733	0.30405321	0.38181143	0.09864803
Dic 2017 ( $j = 2$ )	0.19761239	0.31633172	0.38333545	0.10272044
Dic 2018 ( $j = 3$ )	0.18170888	0.32681106	0.38510538	0.10637468
Dic 2019 ( $j = 4$ )	0.16770487	0.33555746	0.38711248	0.10962519

Tabla 3.3: Predicción para el uso del comercio electrónico desde 2016 a 2019.

A partir de los resultados obtenidos en los ajustes realizados se evidencia una tendencia creciente en la utilización del comercio electrónico para los grupos de edad propuestos en el estudio, sin embargo, para el primer grupo de edad (16 a 44 años) se evidencia un crecimiento mucho mayor en el uso del comercio electrónico frente al segundo grupo de edad (45 a 74 años). Según INE, las razones por la que los consumidores prefieren hacer compras online son la comodidad de este servicio, la posibilidad de encontrar ofertas y artículos a un mejor precio y el ahorro de tiempo que ocasiona no tener que desplazarse físicamente, son cuestiones que se traducen en un aumento significativo para el primer grupo de edad.

### 3.3.3. Ajuste probabilístico

Es necesario encontrar los parámetros del modelo que nos permiten calcular la incertidumbre de los datos (los datos provienen de una encuesta, y por tanto contienen errores aleatorios procedentes del muestreo) y obtener una estimación de la función de densidad de probabilidad de los parámetros. Esta técnica se ha propuesto en (Santonja et al. (2008)) y se ha usado con éxito en (Cortés et al. (2015)). En la Tabla 3.4 se tiene el tamaño de la muestra de cada encuesta.

Asumimos que las salidas de las encuestas son independientes. Para cada una de las 9 encuestas disponibles, denotamos por  $X^j = (X_1^j, X_2^j, X_3^j, X_4^j)$ ,



	Tamaño de la muestra ( $=: n_j$ )
$t_1 = Dic\ 2007\ (j = 1)$	20494
$t_2 = Dic\ 2008\ (j = 2)$	18722
$t_3 = Dic\ 2009\ (j = 3)$	17869
$t_4 = Dic\ 2010\ (j = 4)$	17611
$t_5 = Dic\ 2011\ (j = 5)$	16975
$t_6 = Dic\ 2012\ (j = 6)$	14408
$t_7 = Dic\ 2013\ (j = 7)$	14210
$t_8 = Dic\ 2014\ (j = 8)$	13955
$t_9 = Dic\ 2015\ (j = 9)$	14070

Tabla 3.4: Tamaño de la muestra para cada encuesta considerada.

$0 \leq X_i^j \leq n_j, i = 1, \dots, 4, j = 1, \dots, 9$  un vector aleatorio cuyas entradas son:

- $X_1^j =$  Porcentaje de personas que han usado el comercio electrónico en el primer grupo de edad.
- $X_2^j =$  Porcentaje de personas que no han usado el comercio electrónico en el primer grupo de edad.
- $X_3^j =$  Porcentaje de personas que han usado el comercio electrónico en el segundo grupo de edad.
- $X_4^j =$  Porcentaje de personas que no han usado el comercio electrónico en el segundo grupo de edad.

Estas componentes representan selecciones exclusivas (eventos) con probabilidades dadas por:

- $\mathbb{P}^j(X_1^j = x_1) = \sigma_1^j, \quad j = 1, \dots, 9.$
- $\mathbb{P}^j(X_2^j = x_2) = \sigma_2^j, \quad j = 1, \dots, 9.$

- $\mathbb{P}^j(X_3^j = x_3) = \sigma_3^j, \quad j = 1, \dots, 9.$
- $\mathbb{P}^j(X_4^j = x_4) = \sigma_4^j, \quad j = 1, \dots, 9.$

donde  $\sigma_1^j, \sigma_2^j, \sigma_3^j$  y  $\sigma_4^j$  son los porcentajes recopilados en cada encuesta  $j$  con  $j = 1, \dots, 9$ . Además cada vector aleatorio tiene una distribución de probabilidad multinomial. Así, la probabilidad que  $X_1^j$  ocurra  $x_1$  veces,  $X_2^j$  ocurra  $x_2$  veces,  $X_3^j$  ocurra  $x_3$  veces y  $X_4^j$  ocurra  $x_4$  veces está dado por:

$$\mathbb{P}_{n_j}^j(x_1, x_2, x_3, x_4) = \frac{n_j!}{x_1!x_2!x_3!x_4!} (\sigma_1^j)^{x_1} (\sigma_2^j)^{x_2} (\sigma_3^j)^{x_3} (\sigma_4^j)^{x_4}, \quad j = 1, \dots, 9$$

donde  $x_1 + x_2 + x_3 + x_4 = n_j$ .

Las funciones de probabilidad tetranomiales resultantes para cada encuesta están dadas en la Tabla 3.5.

Año	Distribución de probabilidad asociada				
$t_1 = Dic\ 2007\ (j = 1)$	$\mathbb{P}_{20494}^1 = \frac{20494}{x_1!x_2!x_3!x_4!}$	$0.41544^{x_1}$	$0.1415612^{x_2}$	$0.382352^{x_3}$	$0.040044^{x_4}$
$t_2 = Dic\ 2008\ (j = 2)$	$\mathbb{P}_{18722}^2 = \frac{18722}{x_1!x_2!x_3!x_4!}$	$0.39553^{x_1}$	$0.1790908^{x_2}$	$0.382272^{x_3}$	$0.043105^{x_4}$
$t_3 = Dic\ 2009\ (j = 3)$	$\mathbb{P}_{17869}^3 = \frac{17869}{x_1!x_2!x_3!x_4!}$	$0.36528^{x_1}$	$0.2039887^{x_2}$	$0.375596^{x_3}$	$0.055133^{x_4}$
$t_4 = Dic\ 2010\ (j = 4)$	$\mathbb{P}_{17611}^4 = \frac{17611}{x_1!x_2!x_3!x_4!}$	$0.34256^{x_1}$	$0.2158323^{x_2}$	$0.378195^{x_3}$	$0.060716^{x_4}$
$t_5 = Dic\ 2011\ (j = 5)$	$\mathbb{P}_{16975}^5 = \frac{16975}{x_1!x_2!x_3!x_4!}$	$0.32426^{x_1}$	$0.2284594^{x_2}$	$0.373028^{x_3}$	$0.074248^{x_4}$
$t_6 = Dic\ 2012\ (j = 6)$	$\mathbb{P}_{14408}^6 = \frac{14408}{x_1!x_2!x_3!x_4!}$	$0.28913^{x_1}$	$0.2546902^{x_2}$	$0.371633^{x_3}$	$0.084545^{x_4}$
$t_7 = Dic\ 2013\ (j = 7)$	$\mathbb{P}_{14210}^7 = \frac{14210}{x_1!x_2!x_3!x_4!}$	$0.26688^{x_1}$	$0.2661709^{x_2}$	$0.371846^{x_3}$	$0.095102^{x_4}$
$t_8 = Dic\ 2014\ (j = 8)$	$\mathbb{P}_{13955}^8 = \frac{13955}{x_1!x_2!x_3!x_4!}$	$0.22590^{x_1}$	$0.2958071^{x_2}$	$0.356808^{x_3}$	$0.121487^{x_4}$
$t_9 = Dic\ 2015\ (j = 9)$	$\mathbb{P}_{14070}^9 = \frac{14070}{x_1!x_2!x_3!x_4!}$	$0.18918^{x_1}$	$0.3230407^{x_2}$	$0.345998^{x_3}$	$0.141781^{x_4}$

Tabla 3.5: Función de probabilidad de cada encuesta.

### 3.3.4. Procedimiento del ajuste probabilístico

El procedimiento del ajuste probabilístico viene descrito mediante los siguientes pasos:

- Paso 1: Se consideran los datos del comercio electrónico y su incertidumbre.
- Paso 2: Se muestrean nuevos datos (probables) usando las distribuciones de probabilidad de la Tabla 3.5 (puntos rojos).

- Paso 3: Calibramos los parámetros que hacen que el modelo se ajuste con los datos muestreados (puntos rojos).
- Paso 4: Se repite este procedimiento  $N$  veces, donde  $N$  es un número grande (en nuestro caso  $N = 1000$ ).

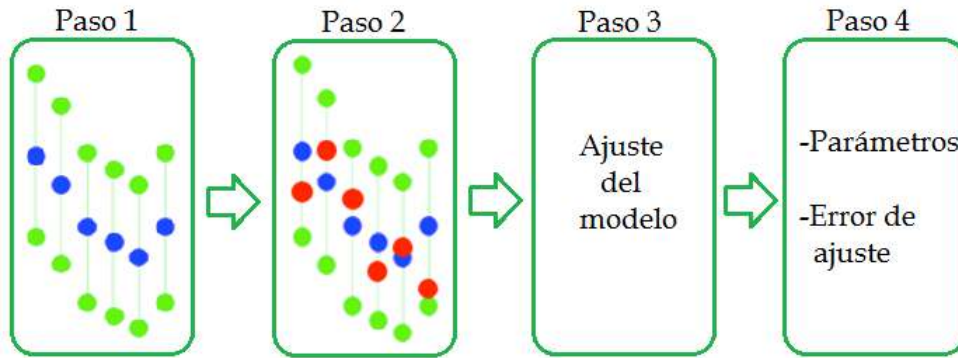


Figura 3.4: Representación gráfica del ajuste probabilístico.

- Paso 5: Ordenamos los parámetros del ajuste del modelo por error (ascendente). Luego tenemos:

$$\{(Param_i, error_i), i = 1, \dots, 1000\} \text{ donde } error_i \leq error_{i+1}.$$

Para seleccionar el conjunto de parámetros que mejor capturan la incertidumbre de los datos del comercio electrónico (representado en la Figura (3.5)), el proceso es el siguiente:

- Tomamos los dos primeros parámetros.
- Los reemplazamos en el modelo.
- Se obtiene la salida del modelo y se calcula la salida del modelo con un intervalo de confianza (DeGroot and Schervish, 2016) del 95 % en cada instante temporal de la Tabla 3.4.

- Se compara los datos del comercio electrónico del intervalo de confianza del 95 % con los de la salida del modelo con un intervalo de confianza del 95 %.
- Se repite con los 3, 4, ..., 1000 primeros parámetros del modelo.

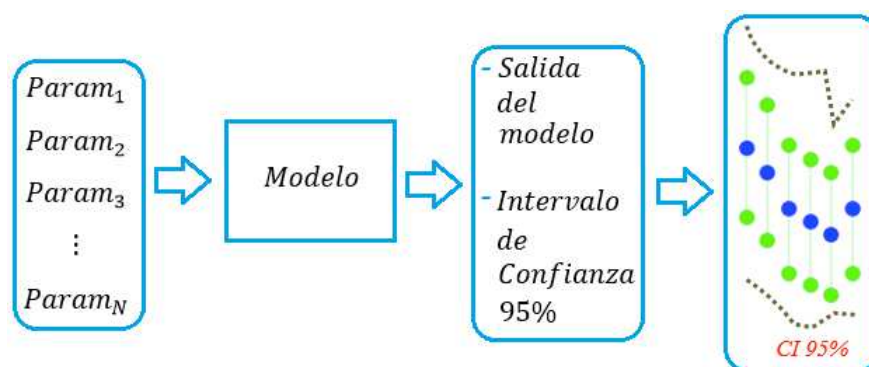


Figura 3.5: Representación gráfica del proceso del ajuste probabilístico.

De esta forma, seleccionamos el que mejor captura la incertidumbre de los datos del comercio electrónico en España y con estos parámetros y el modelo podemos estimar la distribución de probabilidad de los parámetros del modelo y predecir los usuarios del comercio electrónico teniendo en cuenta su incertidumbre.

### 3.3.5. Estimación probabilística

Denotamos  $M(t; Par)$  una abreviada representación de las ecuaciones del modelo (3.10) – (3.13) donde  $Par = (\mu, d_1, c_1, d_2, p_1, \alpha_1, \alpha_2, \gamma_1, p_2, \alpha_3, \alpha_4, \gamma_2)$  son los parámetros del modelo y  $t$  el instante de tiempo en meses.

En la Tabla 3.6 se indican los intervalos de confianza (IC) al 95 % de cada una de las subpoblaciones  $X_i$ ,  $i = 1, 2, 3, 4$ , en cada uno de los años  $t = 2007, \dots, 2015$ . Estos IC se han obtenido a partir de los percentiles 2, 5 y 97,5 después de muestrear a través de la función de probabilidad dada en la Tabla 3.5.

Año	95 % IC de $X_1$	95 % IC de $X_2$	95 % IC de $X_3$	95 % IC de $X_4$
$t_1 = Dic\ 2007$	[0.40840,0.42244]	[0.15629,0.16692]	[0.37581,0.38992]	[0.03728,0.04284]
$t_2 = Dic\ 2008$	[0.38838,0.40271]	[0.17354,0.18484]	[0.37501,0.38956]	[0.04012,0.04611]
$t_3 = Dic\ 2009$	[0.35824,0.37221]	[0.19811,0.20998]	[0.36846,0.38260]	[0.05184,0.05849]
$t_4 = Dic\ 2010$	[0.33814,0.35240]	[0.20978,0.22209]	[0.37090,0.38533]	[0.05708,0.06433]
$t_5 = Dic\ 2011$	[0.31656,0.33190]	[0.22168,0.23542]	[0.36514,0.38097]	[0.07003,0.07867]
$t_6 = Dic\ 2012$	[0.28170,0.29662]	[0.24757,0.26193]	[0.36362,0.37938]	[0.07998,0.08909]
$t_7 = Dic\ 2013$	[0.25966,0.27424]	[0.25876,0.27374]	[0.36367,0.37979]	[0.09029,0.10018]
$t_8 = Dic\ 2014$	[0.21929,0.23262]	[0.28841,0.30320]	[0.34922,0.36462]	[0.11636,0.12679]
$t_9 = Dic\ 2015$	[0.18323,0.19526]	[0.31575,0.33019]	[0.33870,0.35338]	[0.13647,0.14714]

Tabla 3.6: Intervalos de confianza (IC) (95 %) para cada muestra.

Para cada una de las 9 distribuciones de probabilidad tetranomiales dadas en la Tabla 3.5, tomamos una muestra  $d_{j,i}$   $i = 1, \dots, 4$  y  $j = 1, \dots, 9$  y buscamos los valores de los parámetros del modelo  $Par^*$  que mejor ajustan nuestra muestra, que es la que minimiza  $e^*$  dada por:

$$e^* = \left\| \begin{pmatrix} \mathbb{M}(t_1; Par^*) \\ \mathbb{M}(t_2; Par^*) \\ \mathbb{M}(t_3; Par^*) \\ \mathbb{M}(t_4; Par^*) \\ \mathbb{M}(t_5; Par^*) \\ \mathbb{M}(t_6; Par^*) \\ \mathbb{M}(t_7; Par^*) \\ \mathbb{M}(t_8; Par^*) \\ \mathbb{M}(t_9; Par^*) \end{pmatrix} - \begin{pmatrix} d_{1,1} & d_{1,2} & d_{1,3} & d_{1,4} \\ d_{2,1} & d_{2,2} & d_{2,3} & d_{2,4} \\ d_{3,1} & d_{3,2} & d_{3,3} & d_{3,4} \\ d_{4,1} & d_{4,2} & d_{4,3} & d_{4,4} \\ d_{5,1} & d_{5,2} & d_{5,3} & d_{5,4} \\ d_{6,1} & d_{6,2} & d_{6,3} & d_{6,4} \\ d_{7,1} & d_{7,2} & d_{7,3} & d_{7,4} \\ d_{8,1} & d_{8,2} & d_{8,3} & d_{8,4} \\ d_{9,1} & d_{9,2} & d_{9,3} & d_{9,4} \end{pmatrix} \right\|_1, \quad (3.15)$$

donde  $\| \cdot \|_1$  es la 1-norma (Golub and Van Loan, 2012). Este procedimiento es un problema clásico de optimización que ha sido implementado en el algoritmo Nelder-Mead (Nelder and Mead, 1964). Este procedimiento de muestreo-ajuste lo hacemos  $N$  veces, donde  $N$  es un número grande para poder capturar la mayor incertidumbre posible (en nuestro caso,  $N = 1000$ ).

Ahora necesitamos seleccionar sólo los que mejor capturen la incertidumbre de los datos dados por la Tabla 3.6. Con este propósito, nosotros minimizaremos otra función objetivo, esta nos permite determinar sólo las muestras importantes (las que mejor capturan nuestra incertidumbre). Los parámetros de esa función son  $N = 1000$  parámetros booleanos (0 y 1), donde cada uno representa una muestra.

Denotamos como  $\sigma_i, 1 \leq i \leq N$ , si  $\sigma_i = 1$ , la muestra  $i$  es importante para capturar la incertidumbre de los datos, si  $\sigma_i = 0$  esta muestra no se selecciona. De las muestras con  $\sigma_i = 1$ , nosotros calculamos las salidas para los tiempos  $t_1, t_2, \dots, t_9$ . Para cada instante de tiempo ( $t_j$ ), calcularemos los percentiles 2,5 % y 97,5 % para  $n_1(t_j), y_1(t_j), n_2(t_j), y_2(t_j)$ . Por lo tanto, nombramos  $m_1$  la suma de la diferencia, en cada instante de tiempo, entre el percentil 2,5 % de las salidas y el de los datos en la Tabla (3.6). De igual forma, nombramos  $m_2$  la suma de la diferencia entre el percentil 97,5 % de las salidas y el de los datos. Finalmente llamamos  $m = m_1 + m_2$  como el valor que pretendemos minimizar.

### 3.3.6. Resultados

El resultado de la minimización dice que necesitamos  $k = 486$  muestras de las  $N = 1000$  para capturar la incertidumbre con un error de 0,6024. En la Figura 3.6, mostramos para cada grupo  $n_1, y_1, n_2, y_2$  el intervalo de confianza del 95 % y el promedio del modelo ajustado (líneas verdes), por otro lado los puntos rojos representan el intervalo de confianza del 95 % de los datos de la Tabla 3.6. Podemos ver que a pesar de considerar sólo  $k = 486$  muestras, la captura de la incertidumbre es buena. Además en la Figura (3.6) podemos apreciar las predicciones probabilísticas para los próximos cuatro años.

## 3.4. Conclusiones

En este capítulo se ha desarrollado un modelo dinámico de difusión discreto para estudiar las tendencias del comercio electrónico en España.

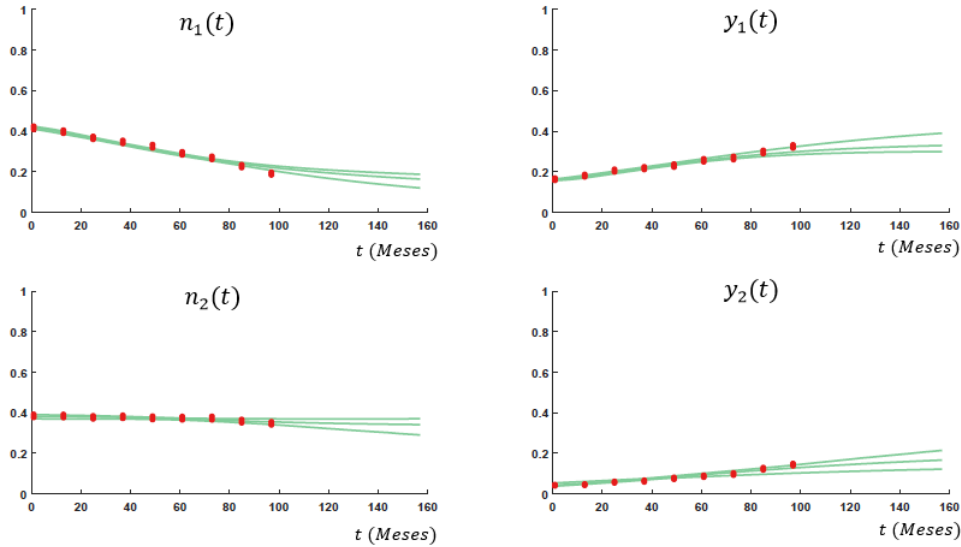


Figura 3.6: Representación gráfica de la incertidumbre del modelo y su ajuste.

Para este estudio se han utilizado datos desde el año 2007 al año 2015 y se clasificaron dos grupos de edad (16 a 44 años y 45 a 74 años) debido a las diferencias significativas que habían frente al uso de esta tecnología. En el estudio se ha supuesto que el tamaño de la población varía en el tiempo. El modelo propuesto se basa en un sistema de ecuaciones en diferencias cuyos parámetros se han determinado implementando un método de optimización denominado PSO. Usando los datos del INE (Instituto Nacional de Estadística en España) se ha ajustado el modelo escalado del comercio electrónico, esto nos ha permitido hacer predicciones para los próximos 4 años. La predicción muestra un aumento sustancial del uso del comercio electrónico, más en el primer grupo de edad que en el segundo.

Debido a que los datos del INE proceden de encuestas que contienen errores de muestreo, para la determinación del ajuste de los parámetros se ha utilizado una técnica de ajuste probabilístico. Este buen ajuste nos permite hacer predicciones probabilísticas fiables sobre el uso del comercio

electrónico en España en los próximos años.



## Capítulo 4

# Modelo de la dinámica de usuarios del comercio electrónico utilizando técnicas de optimización para problemas inversos con incertidumbre

### 4.1. Introducción

El área de las ecuaciones aleatorias se ocupa del estudio, entre otras, de ecuaciones diferenciales ordinarias y ecuaciones en diferencias cuyos datos de entrada (condiciones iniciales/de frontera, y/o coeficientes) se asumen como variables aleatorias ((Dorini et al., 2016), (Hussein and Selim, 2012), (Dorini and Cunha, 2008), (Xu et al., 2016), (Mourad and Debbouche, 2015), (Casabán et al., 2016)). En la mayoría de ellas, se asume que las distribuciones de probabilidad de los parámetros son conocidos y siguen

patrones estándar (Uniforme, Gaussiana, Exponencial, etc.) y el objetivo es ver sus efectos en la solución, la cual es un proceso estocástico. Sin embargo, cuando se plantea modelizar y estudiar un fenómeno real, asumir que la distribución de probabilidad de los parámetros es conocida y estándar no necesariamente es cierto. Este hecho es aún más evidente cuando se trata de un modelo matemático diseñado para describir fenómenos en Ciencias Sociales.

Por lo tanto, determinar las distribuciones de probabilidad apropiadas de los parámetros del modelo es una parte fundamental del problema cuando se trata de modelizar problemas. En otras palabras, cuando se trata de describir fenómenos reales, generalmente no es suficiente construir modelos coherentes, sino también es importante considerar y tratar adecuadamente la incertidumbre involucrada tanto en los datos de las muestras disponibles como en los parámetros del modelo, así como controlar su efecto en la solución. En este último sentido, una cuestión clave es el cálculo de las distribuciones de probabilidad de los parámetros que hacen que la solución del proceso estocástico del modelo capturen la incertidumbre incorporada en los datos de la muestra en instantes de tiempo determinados.

Han habido varias contribuciones donde se aborda este problema. Por ejemplo, en (Casabán et al., 2017, 2016) los autores utilizan las técnicas paramétricas de tipo frecuencialista inversa y de mínimos cuadrados, respectivamente, para describir la incertidumbre de los datos debido a la falta de información sobre cómo se obtuvieron y trataron los datos de la muestra. En ambos enfoques, se asumen implícitamente supuestos restrictivos sobre los parámetros del modelo. En (Cortés et al., 2015), se propone un enfoque computacional completo del problema para obtener la distribución de probabilidad de los parámetros del modelo, el proceso estocástico de la solución y también para capturar la incertidumbre de los datos a través de información puntual (media) y probabilística (intervalos de confianza), pues se requiere asignar previamente las distribuciones de probabilidad de los datos del modelo.

En este capítulo, siguiendo el enfoque presentado en (Cortés et al.,

2015), se pretende capturar la incertidumbre de los datos a través de sus distribuciones de probabilidad en lugar de utilizar solamente la media y los intervalos de confianza. Es decir, dado un modelo y las distribuciones de probabilidad de los datos muestreados que describen su incertidumbre, se desea determinar las distribuciones de probabilidad de los parámetros del modelo de modo que las funciones de densidad de probabilidad del proceso estocástico solución sean similares, en un sentido que se describirá posteriormente, a las funciones de densidad de probabilidad en el instante donde los datos de la muestra estén disponibles.

Existen dos problemas que aumentan la dificultad de este enfoque, por un lado, los propios datos. Las fuentes donde se obtienen los datos no siempre brindan suficiente información para facilitar la asignación de distribuciones de probabilidad que describan su incertidumbre. De todos modos, si se pudieran asignar distribuciones de probabilidad aceptables usando técnicas estadísticas, esto no garantizaría que se describiera con precisión la incertidumbre de todo el fenómeno. Por otro lado, el modelo en sí mismo. Aunque el modelo adquiere flexibilidad cuando sus parámetros se describen mediante distribuciones de probabilidad, se asume implícitamente cierto comportamiento una vez que se apuesta por un modelo (patrón) específico. Por lo tanto, la elección del modelo siempre implica ciertas rigideces que pueden evitar que el proceso estocástico la solución no capture con éxito la incertidumbre de los datos. Aunque la elección de un modelo matemático puede restringir el análisis posterior, a menudo es necesario tomar esta decisión para realizar el estudio correspondiente. De todos modos, esta elección debe basarse en principios que respalden el proceso de modelización, de lo contrario, habría que aplicar métodos heurísticos.

El capítulo está organizado de la siguiente manera. En el Apartado 3.2, se consideran datos sobre los usuarios frecuentes de comercio electrónico en España desde 2011 hasta 2016, y se asignan distribuciones de probabilidad plausibles que capturen la incertidumbre de cada uno de los datos. En la Sección 3.3, se construye un modelo estocástico no lineal para describir la dinámica de los usuarios frecuentes del comercio electrónico en España.

Además, se asignan distribuciones de probabilidad a los parámetros del modelo, discutiendo qué distribuciones deberían ser más flexibles y apropiadas. En el Apartado 3.4 se describen los procedimientos computacionales que se han implementado con el objetivo de determinar qué distribuciones de probabilidad para los parámetros del modelo permitirán que el proceso estocástico de la solución sea lo más similar posible a las funciones de densidad de probabilidad de los datos en los instantes de tiempo correspondientes, usando para este ajuste una técnica de optimización inversa cuya función objetivo involucra medidas de momentos estadísticos. En el Apartado 3.5, se describen detalles relevantes de la implementación y también se muestran los resultados obtenidos. En el Apartado 3.6 está dedicada a la discusión de las principales conclusiones derivadas del presente estudio.

## **4.2. Datos disponibles y modelo del comercio electrónico**

Este apartado está dedicado a presentar los datos disponibles sobre usuarios frecuentes del comercio electrónico que se utilizarán a lo largo del estudio. De acuerdo con el Instituto Nacional de Estadística (INE), los usuarios frecuentes del comercio electrónico son los que han comprado por Internet en los últimos tres meses.

La Tabla 4.1 muestra datos disponibles del INE sobre las proporciones promedio ( $p_k$ ) de personas que usaron el comercio electrónico en España en Dic 2011 ( $k = 0$ ), Dic 2012 ( $k = 1$ ), Dic 2013 ( $k = 2$ ), Dic 2014 ( $k = 3$ ), Dic 2015 ( $k = 4$ ) y Dic 2016 ( $k = 5$ ), así como el tamaño de la muestra ( $n_k$ ) de las encuestas correspondientes. Los instantes de tiempo  $t_k$  están dados en meses (véase Tabla 4.1).

Un punto clave para cuantificar la incertidumbre de los datos que se muestra en la Tabla 4.1 es asignar distribuciones de probabilidad adecuadas de los datos en cada instante  $t_k$ ,  $k = 0, 1, 2, 3, 4, 5$ . Para lograr este objetivo, se usa los tamaños muestrales recogidos en la Tabla 4.1 y se supone que las personas entrevistadas cada año son diferentes y, en consecuencia,

Tabla 4.1: Proporción  $p_k$  de españoles que usaron el comercio electrónico (CE) al menos una vez en los últimos tres meses durante el periodo 2011 – 2016 y el tamaño de muestra  $n_k$  de las encuestas correspondientes (E-commerce, 2018)

Fecha	Instante de tiempo (Meses)	Proporción de usuarios de CE	Tamaño de la muestra (encuesta)
Dic 2011	$t_0 = 0$	$p_0 = 0.186936$	$n_0 = 20647$
Dic 2012	$t_1 = 12$	$p_1 = 0.218087$	$n_1 = 20484$
Dic 2013	$t_2 = 24$	$p_2 = 0.229510$	$n_2 = 20815$
Dic 2014	$t_3 = 36$	$p_3 = 0.276636$	$n_3 = 20786$
Dic 2015	$t_4 = 48$	$p_4 = 0.321732$	$n_4 = 23887$
Dic 2016	$t_5 = 60$	$p_5 = 0.350582$	$n_5 = 24132$

los resultados de las encuestas son independientes. Un individuo puede ser un usuario frecuente del comercio electrónico o no, y estos son dos eventos mutuamente excluyentes. Por lo tanto, es natural considerar que en cada instante de tiempo  $t_k$ ,  $k = 0, 1, 2, 3, 4, 5$ , los usuarios del comercio electrónico se pueden representar mediante una distribución Binomial,  $\text{Bi}(n_k, p_k)$ , donde  $n_k$  es el tamaño de la muestra de encuesta correspondiente y  $p_k \in ]0, 1[$  es la proporción de usuarios frecuentes del comercio electrónico. Para mayor simplicidad, se trabaja con el porcentaje de personas que compran o no en Internet. Por lo tanto, es necesario escalar las distribuciones binomiales al intervalo  $]0, 1[$  y esto se hace dividiendo por el tamaño muestral. En la Tabla 4.2, mostramos las distribuciones binomiales escaladas  $P_k$  para cada instante  $t_k$ ,  $k = 0, 1, 2, 3, 4, 5$  y los momentos de primer y segundo orden. De acuerdo con las propiedades estadísticas de la distribución escalada binomial, los momentos de primer y segundo orden de  $P_k$  se han calculado utilizando las siguientes expresiones

$$m_k = \mathbb{E}[P_k] = p_k, \quad v_k = \mathbb{E}[P_k^2] = \frac{p_k(1-p_k)}{n_k} + p_k^2. \quad (4.1)$$

Tabla 4.2: Distribuciones binomiales escaladas (Bi) para los datos del comercio electrónico de españoles recopilados en la Tabla 4.1 y sus momentos de primer ( $m_k$ ) y segundo orden ( $v_k$ ),  $k = 0, 1, \dots, 5$ , definidos en (4.1).

Instante de tiempo (meses)	Distribución Binomial Escalada	Momento de 1 <sup>er</sup> orden	Momento de 2 <sup>o</sup> orden
$t_0 = 0$	$P_0 = \frac{1}{20647} \text{Bi}(20647, 0.186936)$	$m_0 = 0.186936$	$v_0 = 0.034950$
$t_1 = 12$	$P_1 = \frac{1}{20484} \text{Bi}(20484, 0.218087)$	$m_1 = 0.218087$	$v_1 = 0.047570$
$t_2 = 24$	$P_2 = \frac{1}{20815} \text{Bi}(20815, 0.229510)$	$m_2 = 0.229510$	$v_2 = 0.052683$
$t_3 = 36$	$P_3 = \frac{1}{20786} \text{Bi}(20786, 0.276636)$	$m_3 = 0.276636$	$v_3 = 0.076537$
$t_4 = 48$	$P_4 = \frac{1}{23887} \text{Bi}(23887, 0.321732)$	$m_4 = 0.321732$	$v_4 = 0.103522$
$t_5 = 60$	$P_5 = \frac{1}{24132} \text{Bi}(24132, 0.350582)$	$m_5 = 0.350582$	$v_5 = 0.122917$

### 4.3. Construcción del modelo para estudiar la dinámica de los usuarios frecuentes del comercio electrónico en España

Ahora, se presenta el modelo propuesto para describir la dinámica de los usuarios frecuentes del comercio electrónico en España. En primer lugar se consideran las siguientes subpoblaciones:

- $X_t^{\bar{B}}$  representa el porcentaje o proporción de personas que no utilizan frecuentemente el comercio electrónico en el instante de tiempo  $t$  (en meses),
- $X_t^B$  representa el porcentaje o proporción de personas que utilizan frecuentemente el comercio electrónico en el instante de tiempo  $t$  (en meses).

Por lo tanto,  $X_t^{\bar{B}}, X_t^B \in ]0, 1[$ , para cada  $t$ . En el periodo 2011 – 2016, según (?) (2015), las tasas de natalidad y mortalidad son muy cercanas y, en consecuencia, se asume una población constante. Además, se considera que quienes no son usuarios del comercio electrónico pueden convertirse en usuarios por la influencia de los usuarios del comercio electrónico (Christakis and Fowler, 2009), y esto se modeliza por el término  $\beta X_t^{\bar{B}} X_t^B$ , donde  $\beta > 0$  es el parámetro de transmisión. Las personas pueden dejar de ser usuarios del comercio electrónico ( $X_t^B$ ), cuando no compran ningún artículo por Internet en tres meses, a una tasa  $\gamma > 0$ , y se modeliza con el término  $\gamma X_t^B$ . La dinámica de las personas que usan y no usan el comercio electrónico se describe mediante el siguiente sistema no lineal de ecuaciones en diferencias:

$$\begin{cases} X_{t+1}^{\bar{B}} = X_t^{\bar{B}} - \beta X_t^{\bar{B}} X_t^B + \gamma X_t^B, \\ X_{t+1}^B = X_t^B + \beta X_t^{\bar{B}} X_t^B - \gamma X_t^B. \end{cases} \quad (4.2)$$

En la Figura 4.1, se muestra el diagrama correspondiente al modelo dinámico propuesto.

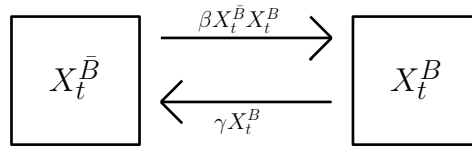


Figura 4.1: Modelo compartimental para la dinámica de la proporción de usuarios frecuentes del comercio electrónico en España.

Como  $X_t^{\bar{B}}$  y  $X_t^B$  son probabilidades de dos eventos mutuamente excluyentes, entonces  $X_t^{\bar{B}} + X_t^B = 1$ , para cada  $t$ . Por lo tanto, el modelo dinámico (4.2) puede reescribirse como:

$$\begin{cases} X_{t+1}^{\bar{B}} = 1 - X_{t+1}^B, \\ X_{t+1}^B = X_t^B + \beta X_t^{\bar{B}} (1 - X_t^B) - \gamma X_t^B. \end{cases} \quad (4.3)$$

Una vez que se ha formulado y simplificado el modelo propuesto, teniendo en cuenta los errores de muestreo en las encuestas, es natural considerar los

parámetros del modelo como variables aleatorias para describir la incertidumbre de los datos. En una primera etapa, se considera el parámetro  $\beta$  como una variable aleatoria positiva y  $\gamma$  como una constante positiva que debe ser determinada.

En (Brauer and Castillo-Chávez, 2001), los autores muestran que una distribución de probabilidad razonable para el parámetro  $\gamma$  es la exponencial. En las pruebas y simulaciones realizadas por computadora, se encuentra que un pequeño porcentaje de usuarios del comercio electrónico se mueve al estado de no usuarios, lo que concuerda con las observaciones recopiladas en (E-commerce, 2018)(2015). Esto significa que la media, y en consecuencia la varianza, de la distribución exponencial son muy pequeñas y admite que se pueda considerar el parámetro del modelo  $\gamma$  como una constante determinística, en lugar de una variable aleatoria, a determinar.

El parámetro de transmisión  $\beta$  incorpora procesos sociales complejos que implican el hábito frecuente del uso del comercio electrónico. Estos procesos sociales generalmente no se estudian en encuestas estándar, como ocurre en (E-commerce, 2018)(2015). Por lo tanto, es plausible considerar  $\beta$  como una variable aleatoria. Por ello asumiremos que la variable aleatoria  $\beta$  sigue una distribución Beta prime  $Be_p(\beta_1, \beta_2)$  ya que es positiva y flexible (biparamétrica) (Norman et al., 1994).

De (Brauer and Castillo-Chávez, 2001, pag. 352), el parámetro  $\gamma$  se puede interpretar como el tiempo promedio que un usuario del comercio electrónico se convierte en un no usuario. Teniendo en cuenta que las encuestas consideran que un usuario frecuente del comercio electrónico cambia su estado actual si no compra por Internet durante tres meses y el paso de tiempo del modelo es un mes, el parámetro del modelo  $\gamma$  estará acotado superiormente por  $1/3$  y debe estar en el intervalo  $(0, 1/3)$ .

Ahora, el objetivo es buscar los parámetros  $\beta_1 > 0$  y  $\beta_2 > 0$  de la variable aleatoria positiva  $\beta$ , y el número real  $\gamma > 0$ , de modo que la salida del modelo formulado en (4.3) se ajustan lo mejor posible a las distribuciones de probabilidad en los instantes de tiempo correspondientes que se muestran en la Tabla 4.2, según algún criterio de error que se establezca.



#### 4.4. Procedimiento para resolver el problema utilizando una técnica de optimización inversa

Se propone un algoritmo computacional para encontrar los valores de los parámetros  $\beta_1$ ,  $\beta_2$  y  $\gamma$  tales que las distribuciones de probabilidad del modelo (4.3) en los instantes de tiempo  $t_0, t_1, t_2, t_3, t_4, t_5$ , estén lo más cerca posible de las correspondientes distribuciones de probabilidad de la Tabla 4.2. Para lograr este objetivo, se propone un método computacional para minimizar la función de error del problema no lineal inverso asociado. En primer lugar, según la segunda ecuación de (4.3), definimos la función

$$M(X, \beta, \gamma) = X + \beta X(1 - X) - \gamma X, \quad (4.4)$$

eso permite calcular los usuarios frecuentes del comercio electrónico en un mes dependiendo de los usuarios frecuentes del comercio electrónico en el mes anterior. Un paso importante en nuestro procedimiento es obtener el resultado del modelo cada mes dadas las estimaciones de los parámetros del modelo. Esto se realizará mediante la función de salida, dada por  $O(\beta_1, \beta_2, \gamma, T)$ , donde  $T$  denota la cantidad de meses para calcular la salida del modelo  $mo$ , definida por los siguientes pasos:

- Inicializar el vector  $mo$  con el tamaño  $T + 1$ .
- Sea  $x_0$  el valor de la condición inicial muestreada de acuerdo con la distribución binomial escalada  $P_0$  (véase Tabla 4.2).
- Sea  $b$  un valor muestreado de acuerdo con la distribución Beta Prima  $\beta = \text{Be}_p(\beta_1, \beta_2)$ .
- Asigna  $mo(1) = x_0$ .
- Para  $i = 1$  hasta  $T$ 
  - Calcular  $mo(i + 1) = M(mo(i), b, \gamma)$ .
- Devolver  $mo$ .

Tendremos que repetir varias veces la evaluación de la función anterior  $O(\beta_1, \beta_2, \gamma, T)$ . Por lo tanto, definimos una función que realiza esta tarea. Definamos la función  $S(\beta_1, \beta_2, \gamma, N, T)$ , donde  $N$  es el número de veces que evaluamos la función  $O$ , definida por:

- Inicializar la matriz  $OUT$  de tamaño  $N \times T$ .
- Para  $i = 1$  hasta  $N$ 
  - Calcular la fila  $OUT(i) = O(\beta_1, \beta_2, \gamma, T)$ .
- Calcular  $OUT$ .

El algoritmo anterior devuelve una matriz de tamaño  $N \times T$ , donde cada fila es el vector objetivo (salida del modelo) obtenida con un valor de la muestra de  $x_0$  y  $\beta$ .

Ahora, presentamos la función de ajuste que nos ayudará a determinar los mejores parámetros del modelo  $\beta_1$ ,  $\beta_2$  y  $\gamma$ . La función de ajuste aplica el algoritmo anterior para obtener una matriz  $OUT$  con  $N$  ejecuciones. Sin embargo, no necesitamos todos los elementos de la matriz  $OUT$ , necesitamos todas las filas, pero solo en los instantes de tiempo  $t_k$ , para  $k$  de 0 a 5, donde tenemos datos y luego podremos comparar. Se define la función de ajuste  $F(\beta_1, \beta_2, \gamma)$  de la siguiente manera:

- Sea  $N$  un valor fijo, lo suficientemente grande.
- Sea  $T = 60$ .
- Calcular la matriz  $OUT = S(\beta_1, \beta_2, \gamma, N, T)$ .
- Para  $k = 0$  hasta 5
  - Tomar la columna  $t_k$  de la matriz  $OUT$ , es decir, las  $N$  salidas del modelo correspondientes al mismo instante de tiempo  $t_k$   $\{OUT(1, t_k), \dots, OUT(N, t_k)\}$  y calcular  $M_k$ , su media (momento del primer orden) y  $V_k$ , su momento de segundo orden.

- Calcular

$$E = |M_0 - m_0| + \dots + |M_5 - m_5| + |V_0 - v_0| + \dots + |V_5 - v_5|.$$

- Obtener  $E$ .

La función objetivo  $F(\beta_1, \beta_2, \gamma)$  es una función con valores reales positivos definida en el dominio  $\mathbb{R}^+ \times \mathbb{R}^+ \times (0, 1/3)$ . El dominio inicial es demasiado grande y, después de algunas ejecuciones, lo reduciremos.

Ahora, para encontrar los valores apropiados de  $\beta_1$ ,  $\beta_2$  y  $\gamma$ , debemos minimizar la función objetivo  $F$ . Para hacer eso, utilizamos el algoritmo de optimización de enjambre de partículas aleatorias (rPSO) presentado en (Khemka and Jacob (2010)). Este algoritmo de optimización es una variación del algoritmo PSO clásico donde se introduce la aleatoriedad en el cálculo de la nueva velocidad de las partículas. rPSO ha sido utilizado con éxito para ajustar los parámetros de los modelos con incertidumbre de los datos (Acedo et al., 2017).

Por tanto, aplicamos el procedimiento anterior para ajustar de manera probabilística el modelo propuesto a las distribuciones de los datos de la Tabla 4.2. Ahora, describimos algunas características adicionales incluidas en nuestra versión del algoritmo rPSO:

1. Incluimos un 10% de probabilidad de mutación.
2. Aplicamos rPSO con 90 partículas. 60 de ellas siguen el procedimiento clásico de actualización de PSO. El resto (30), llamadas *exploradoras*, se generan actualizaciones que se almacenan en un repositorio. Por lo tanto, la cantidad de exploradoras almacenados en el repositorio aumenta a medida que pasa el tiempo. En cada iteración (una iteración corresponde a la evaluación de una generación completa de partículas), calculamos el centro de masas de una muestra de 1/3 de las partículas en el repositorio y actualizamos las velocidades de las exploradoras tomando direcciones lejanas del centro de masas.

Observe que las características anteriores nos permitirán explorar más extensamente el espacio de los parámetros.

## 4.5. Resultados

Para las evaluaciones de la función objetivo  $F(\beta_1, \beta_2, \gamma)$ , hemos decidido tomar  $N = 1000$  y el espacio de búsqueda de parámetros limitado a  $\beta_1 \in [1000, 10000]$ ,  $\beta_2 \in [50000, 150000]$  y  $\gamma \in (0, 1/3)$ . El intervalo para  $\gamma$  se justificó previamente. Los intervalos para  $\beta_1$  y  $\beta_2$  se han elegido después de algunas ejecuciones y teniendo en cuenta que la distribución Beta Prime tiene una gran variación con pequeñas variaciones de sus parámetros.

Respecto al algoritmo rPSO, se consideraron 90 partículas con un número máximo de iteraciones de 5000. El procedimiento se ejecutó en un PC con CPU Intel i5 y 16 GB de RAM en MS-Windows 10. El código se implementó en Python 2.7 y el procedimiento de ajuste duró 1 hora y 52 minutos.

Los valores de los parámetros con mejor ajuste fueron:

$$\beta_1 = 1985,7534, \beta_2 = 13820,0903, \gamma = 7,02012261 \times 10^{-5}, \quad (4.5)$$

siendo el error cometido de:

$$F(1985,7534, 13820,0903, 7,02012261 \times 10^{-5}) = 0,0435104687349.$$

Gráficamente, el resultado del ajuste se puede ver en la Figura 4.2. Aquí se evalúa la función  $S(\beta_1, \beta_2, \gamma, 1000, 60)$  con  $\beta_1, \beta_2, \gamma$  los valores dados en (4.5), y creamos los histogramas de los valores de salida del modelo  $N = 1000$  en los instantes de tiempo  $t_k$ ,  $\{OUT(1, t_k), \dots, OUT(N, t_k)\}$ , para  $t_k = 12, 24, 36, 48, 60$  meses (en azul). Comparamos estos histogramas con los histogramas de las distribuciones de probabilidad de los datos  $P_1, P_2, P_3, P_4, P_5$  de la Tabla 4.2 (en rojo). A excepción de diciembre de 2013, la similitud de los histogramas en cada instante de tiempo es notable porque ambos histogramas comparten la mayor parte de sus incertidumbres.

En la Figura 4.3, podemos ver la función de densidad de probabilidad obtenida del parámetro de transmisión  $\beta$ , la distribución Beta Prima de la variable aleatoria  $\beta = \text{Be}_p(1985,7534, 13820,0903)$ . Como el ajuste ha sido

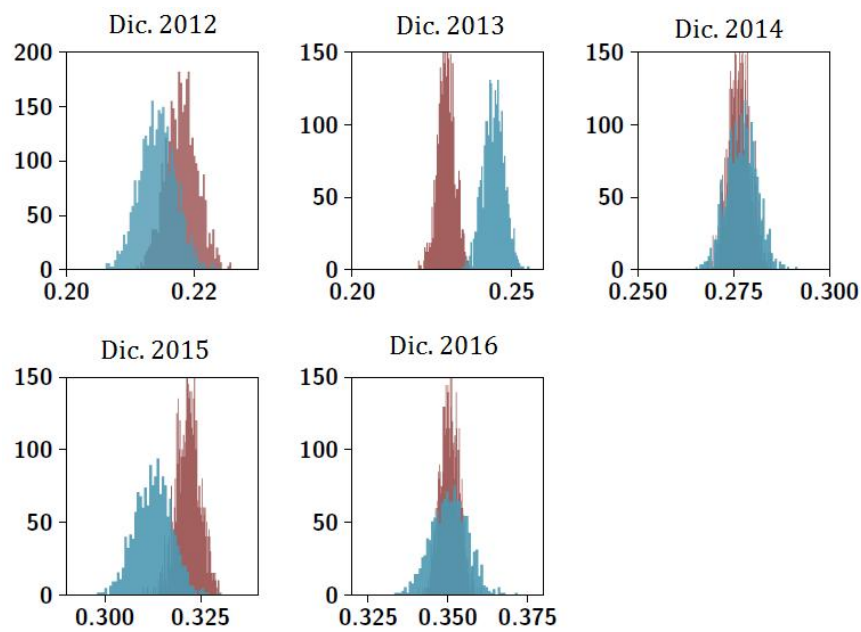


Figura 4.2: Comparación entre los histogramas de los resultados del modelo  $N = 1000$  para  $t_1 = \text{Dic } 2012$ ,  $t_2 = \text{Dic } 2013$ ,  $t_3 = \text{Dic } 2014$ ,  $t_4 = \text{Dic } 2015$ ,  $t_5 = \text{Dic } 2016$  (en azul) y sus histogramas de datos de las distribuciones binomiales escaladas.

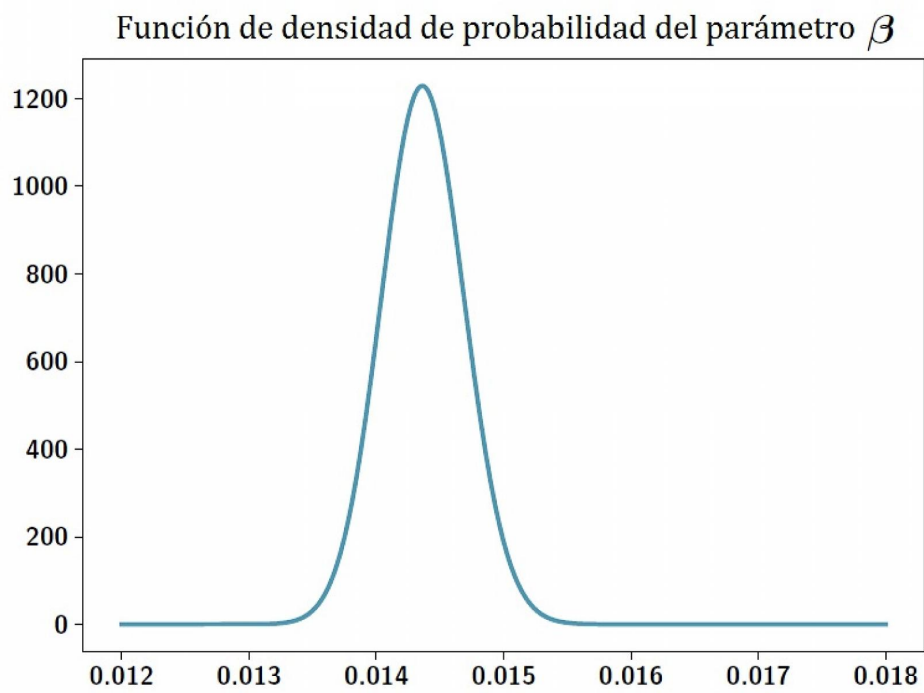


Figura 4.3: Función de densidad de probabilidad del parámetro de transmisión de variable aleatoria  $\beta = \text{Be}_p(\beta_1, \beta_2)$ .

bueno, podemos evaluar  $S(\beta_1, \beta_2, \gamma, 1000, 108)$  con  $\beta_1, \beta_2, \gamma$  dados en (4.5) y luego los histogramas se construyen a partir de los  $N = 1000$  valores de salida del modelo en los instantes de tiempo  $t_k$ ,  $\{OUT(1, t_k), \dots, OUT(N, t_k)\}$ , para  $t_k = 72, 84, 96, 108$ , para realizar predicciones del modelo mediante las distribuciones de probabilidad de los usuarios frecuentes del comercio electrónico en Dic de 2017, Dic de 2018, Dic de 2019 y Dic de 2020, respectivamente. Tomando  $T = 108$  meses, llegamos a diciembre de 2020. Los resultados gráficos se ilustran en la Figura 4.4.

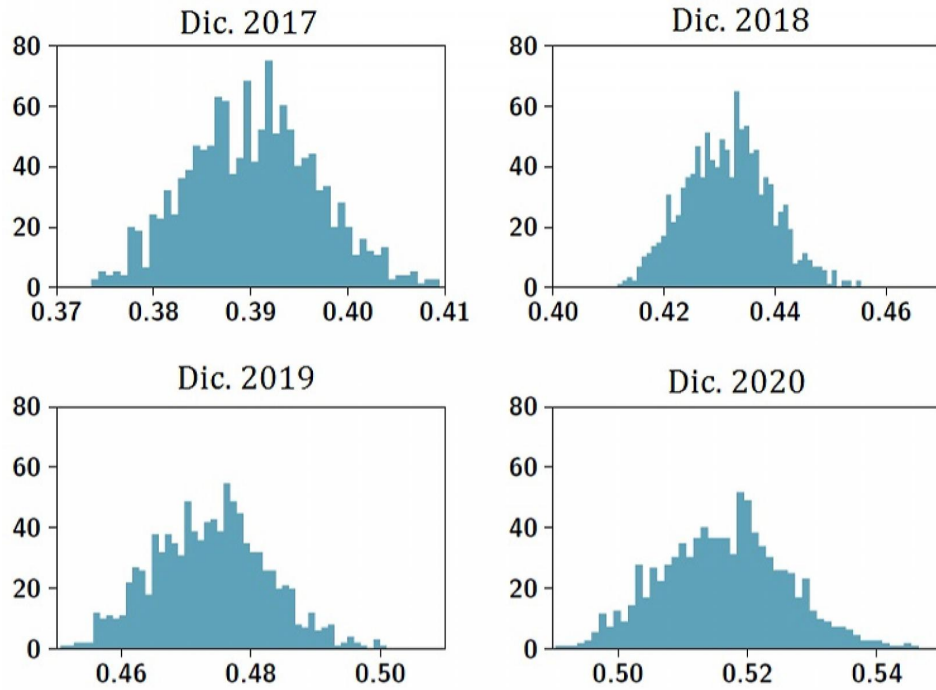


Figura 4.4: Histogramas de los resultados del modelo  $N = 1000$  en los instantes de tiempo Dic 2017, Dic 2018, Dic 2019 y Dic 2020.

En la Figura 4.5, mostramos los intervalos de confianza (DeGroot and Schervish, 2016) del 95 % de la distribución de datos (variables aleatorias

binomiales escaladas en la Tabla 4.2) representadas por los puntos rojos, y el modelo con intervalos de confianza del 95 % representados por las líneas verdes en cada mes desde 2011–2016 (estimación) y desde 2017–2020 (predicción). Esta es otra forma de visualizar la bondad de ajuste del modelo y la predicción.

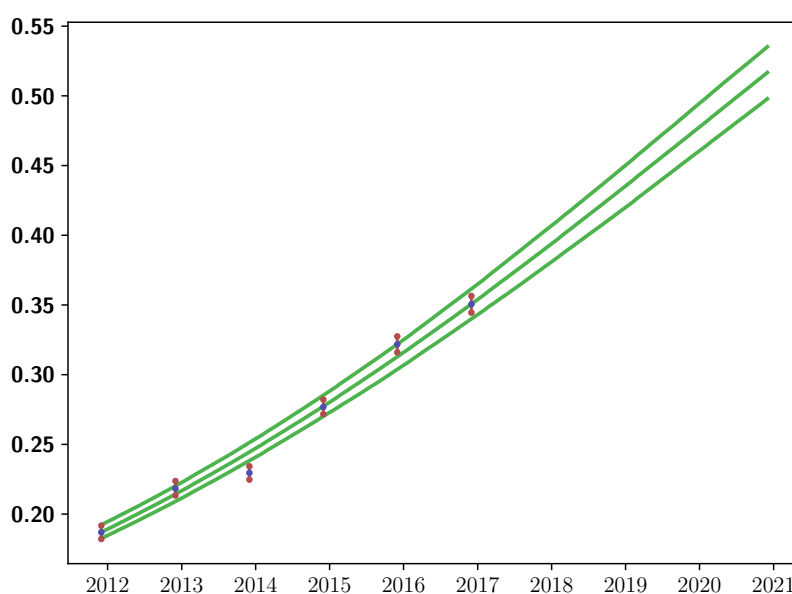


Figura 4.5: La media y el intervalo de confianza del 95 % representados por puntos rojos y azules para los datos de Dic. de 2011 a Dic. de 2016. En el eje  $X$  denotamos los años y en el eje  $Y$  los porcentajes. Las líneas verdes representan el promedio e intervalo de confianza del 95 % para los meses desde Dic. de 2011 hasta Dic. 2020.

## 4.6. Conclusiones

En este capítulo se ha propuesto un modelo dinámico discreto sencillo para describir la evolución del comercio electrónico en España. La particu-



laridad de este modelo, frente a los propuestos en los capítulos anteriores, es que el coeficiente de difusión de esta tecnología se trata como una variable aleatoria y se le asigna una distribución adecuada (en nuestro caso una Beta Prime), la cual depende de dos parámetros. En el capítulo se ha propuesto un método computacional para estimar dichos parámetros (junto a otros parámetros de naturaleza determinista de los cuales también depende el modelo propuesto) de modo que la respuesta estocástica del modelo sea capaz de ajustar los datos reales (que sobre el uso del comercio electrónico en España proporciona el INE), y de capturar la incertidumbre que contiene la información recogida en las muestras del INE. El resultado es un método computacional bastante general y flexible, en el sentido que puede adaptarse a otros modelos estocásticos, que en el caso del comercio electrónico consigue capturar el comportamiento de los datos reales y su variabilidad. Las medidas de bondad de ajuste calculadas permiten validar el modelo y aplicarlo para hacer predicciones reales en el corto plazo. Estas predicciones se completan con información dada a través de histogramas tanto el uso del comercio electrónico en los próximos años, como para el coeficiente de difusión de esta tecnología, ya que el modelo se trata desde el punto de vista estocástico.



## Capítulo 5

# Conclusiones

El trabajo presentado en la tesis muestra un estudio sobre el uso del comercio electrónico en España y sus tendencias a partir de datos reales tomados del Instituto Nacional de Estadística (INE). En primer lugar, se realiza un modelo determinista para una población objetivo de 16 a 74 años. Esta población, distribuida en 6 grupos de edad sobre quienes hacían uso del comercio electrónico o no, fue modelizada con incertidumbre utilizando técnicas de epidemiología matemática. Para esto, fue elaborado un modelo demográfico con datos del INE y con esta información, un sistema no lineal de ecuaciones diferenciales fue construido para analizar las tendencias del comercio electrónico por grupos de edades. Se implementan técnicas computacionales que permitan ajustar el modelo utilizando el algoritmo de Nelder-Mead y hallar los valores de los parámetros que proporcionan la mejor solución. Para el enfoque con incertidumbre se implementa una técnica computacional a partir del LHS (Latin Hypercube Sampling) y de esta forma realizar un análisis de sensibilidad y estabilidad de los parámetros ajustados. Algunos de los limitantes para este estudio fueron los pocos datos frente al estudio propuesto, la forma en que se divide cada grupo de edad y los resultados del método determinista de optimización implementado, ya que partir de condiciones iniciales dadas en el modelo, al realizar varias ejecuciones siempre encontrará el mismo mínimo o máximo global.

En busca del mejoramiento de la técnica del estudio de la incertidumbre para el problema propuesto, en el Capítulo 2 se amplía el estudio, se dispone de más datos sobre el E-commerce en España. En este caso, la población objetivo se divide en dos grupos de edad, debido a resultados significativos del uso del E-commerce en estos grupos de edad. En la modelización de los grupos de edad se utilizan técnicas de epidemiología matemática, a partir de un sistemas de ecuaciones en diferencias para predecir el uso del E-commerce en la población objetivo. El método computacional implementado para analizar la incertidumbre de los datos fue PSO (Particle Swarm Optimization) que mejora los resultados obtenidos en el estudio anterior. Este algoritmo es un método heurístico que tiene cierta aleatoriedad dentro de la búsqueda de un mínimo o máximo global y esto hace que a pesar de encontrar un mínimo o máximo global, por su aleatoriedad pueda encontrar una ruta que mejore el mínimo o máximo global encontrado anteriormente. Para mejorar el análisis de la incertidumbre de los datos planteados, en el Capítulo 3 se construye un modelo computacional usando una variación del algoritmo de enjambre de partículas (PSO) que determine la distribución de probabilidad que se puede asignar a los parámetros del modelo, y de esta forma, se implemente un proceso estocástico que capture la incertidumbre de los datos a través de distribuciones de probabilidad. Las predicciones para los usuarios del E-commerce en España fue realizada a partir de histogramas. Para esto se muestreaba en la distribución de probabilidad, estos valores los sustituimos en el modelo y se obtenían las salidas del modelo, dados en histogramas. Esto nos permite hacer predicciones confiables durante los próximos cuatro años.

El trabajo muestra la construcción de modelos deterministas de un fenómeno real a partir de datos oficiales. El estudio de la incertidumbre de los datos y la construcción de modelos estocásticos muestran una evolución en las técnicas computacionales implementadas para el estudio de la incertidumbre para hacer pronósticos en el corto plazo del E-commerce

en España. Esto permite mostrar un aumento progresivo en la implementación del E-commerce al momento de realizar compras o transacciones, lo que sugiere a las empresas invertir más en el desarrollo de infraestructuras digitales seguras, estrategias de marketing para consolidar a los consumidores existentes y llegar a los consumidores potenciales debido al aumento en la población internauta.



# Bibliografía

- Acedo, L., Burgos, C., Hidalgo, J.-I., Sánchez-Alonso, V., Villanueva, R.-J., and Villanueva-Oller, J. (2017). Calibrating a large network model describing the transmission dynamics of the human papillomavirus using a particle swarm optimization algorithm in a distributed computing environment. *The International Journal of High Performance Computing Applications*, pages 721–728.
- Blower, S. M. and Hens, D. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: and hiv model, as an example. *International Statistical Review*, 62(2):229–243.
- Brauer, F. and Castillo-Chávez, C. (2001). *Mathematical Models in Population Biology and Epidemiology*. Springer New York.
- Casabán, M.-C., Cortés, J.-C., Navarro-Quiles, A., Romero, J.-V., Roselló, M.-D., and Villanueva, R.-J. (2016). A comprehensive probabilistic solution of random SIS-type epidemiological models using the random variable transformation technique. *Communications in Nonlinear Science and Numerical Simulation*, 32:199–210.
- Casabán, M.-C., Cortés, J.-C., Navarro-Quiles, A., Romero, J.-V., Roselló, M.-D., and Villanueva, R.-J. (2017). Computing probabilistic solutions of the Bernoulli random differential equation. *Journal of Computational and Applied Mathematics*, 309:396–407.
- Christakis, N. A. and Fowler, J. H. (2009). *Connected. The surprising power*

*of our social networks and how they shape our lives.* Little, Brown & Company.

Cortés, J.-C., Sánchez, F., Santonja, F.-J., and Villanueva, R.-J. (2015). A probabilistic analysis to quantify the effect of march 11, 2004, attacks in madrid on the march 14 elections in spain: a dynamic modelling approach. In *Abstract and Applied Analysis*, volume 2015. Hindawi.

Cortés, J.-C., Santonja, F.-J., Tarazona, A.-C., Villanueva, R.-J., and Villanueva-Oller, J. (2015). A probabilistic estimation and prediction technique for dynamic continuous social science models: The evolution of the attitude of the basque country population towards ETA as a case study. *Applied Mathematics and Computation*, 264:13–20.

DeGroot, M. H. and Schervish, M. J. (2016). *Probability & Statistics*, volume 1. Pearson.

Dorini, F. and Cunha, M. (2008). Statistical moments of the random linear transport equation. *Journal of Computational Physics*, 227(19):8541–8550.

Dorini, F. A., Cecconello, M. S., and Dorini, M. B. (2016). On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density. *Communications in Nonlinear Science and Numerical Simulation*, 33:160–173.

E-commerce, E. S. E. (2018). <http://www.ine.es>.

Eberhart, R. and Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, pages 39–43. IEEE.

Frambach, R. T. (1993). An integrated model of organizational adoption and diffusion of innovations. *European Journal of Marketing*, 27:22–41.

Golub, G. H. and Van Loan, C. F. (2012). *Matrix computations*, volume 3. JHU press.



- Hethcote, H. W. (2000). The mathematics of infectious diseases. *Society For Industrial and Applied Mathematics*, 42:599–653.
- Hoare, A., D.G., R., and D.P., W. (2008). Sampling and sensitivity analysis tools (sasat) for computational modeling. *Theoretical Biology and Medical Modeling*, 5(4).
- Hussein, A. and Selim, M. M. (2012). Solution of the stochastic radiative transfer equation with Rayleigh scattering using RVT technique. *Applied Mathematics and Computation*, 218(13):7193–7203.
- INE (2008). <http://www.ine.es>.
- INE (2015a). <https://www.ine.es/jaxi/Tabla.htm?path=/t25/p450/base2011>.
- INE (2015b). <http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Esta>.
- Jacob, C. and Khemka, N. (2004). Particle swarm optimization in mathematica an exploration kit for evolutionary optimization. In *Proceedings of the Sixth International Mathematica Symposium*.
- Khemka, N. and Jacob, C. (2010). Exploratory toolkit for evolutionary and swarm-based optimization. *The Mathematica Journal*, 11(3):376–391.
- Lombana, I.-C., Santonja, F.-J., , Rubio, M., Sánchez, E., and Villanueva, J. (2010). A network model for the short-term prediction of the evolution of cocaine consumption in spain. *Mathematical and Computer Modelling*, 52(7-8):1023–1029.
- Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254:178–196.
- Martcheva, M. and Castillo-Chavez, C. (2003). Diseases with chronic stage in a population with varying size. *Mathematical Biosciences*, 182(1):1–25.
- Mas, F. G. (2004). *Comercio y firma electrónicos (Análisis jurídico de los servicios de la sociedad de la información)*. Editorial Lex Nova.

Mathematica (2008). <http://www.wolfram.com/products/mathematica>.

Mena-Lorca, J. and Hethcote, H. W. (1992). Dynamic models of infectious diseases as regulators of population sizes. *Journal of mathematical biology*, 30(7):693–716.

Mourad, K. and Debbouche, A. (2015). Complete controllability of non-local fractional stochastic differential evolution equations with Poisson jumps in Hilbert spaces. *International Journal of Advances in Applied Mathematics and Mechanics*, 3(1):41–48.

Nelder, J. A. and Mead, R. (1964). A simplex method for function minimization. *The Computer Journal*, 7:308–313.

Norman, L., Kotz, S., and Balakrishnan, N. (1994). Continuous univariate distributions.

Olsson, A., G., S., and O., D. (2003). On latin hypercube sampling for structural reliability analysis. *Structural Safety*, 25:47–68.

Owen, A. (1994). Controlling correlations in latin hypercube samples. *Journal of American Statistical Association*, 89(428):1517–1522.

Roger, E. M. and Olaguera, N. S. (2003). *Diffusion of innovations*. Simon and Schuster Adult Publishing Group.

Santonja, F. J., Tarazona, A. C., and Villanueva, R. J. (2008). A mathematical model of the pressure of an extreme ideology on a society. *Computers & Mathematics with Applications*, 56(3):836–846.

Völlink, T., Meertens, R. M., and Midden, C. J. (2006). Diffusion of technological innovations. In *User Behavior and Technology Development*, pages 173–180. Springer.

Xu, Z., Tipireddy, R., and Lin, G. (2016). Analytical approximation and numerical studies of one-dimensional elliptic equation with random coefficients. *Applied Mathematical Modelling*, 40(9–10):5542–5559.

- Zhang, D. and Ntoko, A. (2002). Mathematical model of technology diffusion in developing countries. *Computational Methods in Decision-Making, Economics and Finance*, pages 526–539.

