



Fac. CC. Económicas y Empresariales
Universidad de La Laguna



Fac. CC. Económicas y Empresariales
Univ. de Las Palmas de Gran Canaria

Predicción mediante algoritmos genéticos con matriz de transición. Una aplicación a la demanda turística de Tenerife

Montserrat Hernández López *

DOCUMENTO DE TRABAJO 2005-02

* Universidad de La Laguna. Departamento de Economía de las Instituciones, Estadística Económica y Econometría.

**PREDICCIÓN MEDIANTE ALGORITMOS GENÉTICOS CON MATRIZ DE
TRANSICIÓN. UNA APLICACIÓN A LA DEMANDA TURÍSTICA EN
TENERIFE**

MONTSERRAT HERNÁNDEZ LÓPEZ¹

Dpto. de Economía de las Instituciones, Estadística Económica y Econometría

Facultad de Ciencias Económicas y Empresariales

Universidad de La Laguna

Campus de Guajara. 38071 La Laguna

Tenerife. Islas Canarias. España

Teléfono: 922317032

e-mail: mhdezl@ull.es

¹ La autora desea expresar su agradecimiento sincero al evaluador anónimo de la Colección de Documentos de Trabajo ULL-ULPGC sus útiles sugerencias que, sin duda, han contribuido a mejorar notablemente la versión final de este documento.

RESUMEN

Aunque los algoritmos genéticos se diseñaron originalmente como método de optimización, también pueden ser utilizados, en el contexto económico, como herramienta predictiva de los cambios en la composición de una población, en términos de las características individuales de los agentes que la componen. En este documento, se desarrolla un algoritmo genético específico capaz de predecir los cambios en las características de los turistas que visitan el sur de Tenerife. Los resultados obtenidos sugieren la conveniencia de sustituir los tradicionales operadores de cruce y mutación por la acción de una adecuada matriz de transición. Esta matriz dirige la dinámica de transformación de la población en el sentido de que permite introducir consideraciones económicas que otorgan mayores probabilidades a ciertas transformaciones en las características de los turistas que visitan la Isla.

PALABRAS CLAVE. Algoritmos genéticos, predicción económica, operadores genéticos, matriz de transición.

1. INTRODUCCIÓN

Los algoritmos genéticos describen el proceso de transformación de los individuos de una población, definidos mediante un vector de características, de acuerdo con fundamentos teóricos (Holland, 1975; Goldberg, 1989) que se basan en los procesos genéticos de formación, aprendizaje, adaptación y evolución de los organismos biológicos, y, especialmente, en el principio de selección natural o *supervivencia de los mejores* —término acuñado por Charles Darwin en *The Origin of Species* de 1859—, y en los resultados de los intercambios genéticos. En la naturaleza, los procesos de selección natural y alteración genética ocurren en una generación, y luego en su descendencia, y a continuación en la descendencia de ésta, y así sucesivamente. Después de cada generación, o al menos así se desea, la población será mejor —o más adaptada a su entorno— que las anteriores, es decir, los individuos estarán más evolucionados (Moreno y Moreno, 1999).

Tradicionalmente, los algoritmos genéticos han sido considerados y aplicados como algoritmos de búsqueda de soluciones mejores —no necesariamente óptimas—, especialmente útiles en problemas específicos de gran dimensión. Esta técnica difiere de los procedimientos tradicionales de optimización en varios aspectos que contribuyen a la robustez del algoritmo genético en cierto tipo de problemas y le proporcionan una ventaja sobre otras técnicas de optimización². Pero, además de su uso con fines de optimización, los algoritmos genéticos podrían considerarse como radiografías del proceso dinámico de transformación interna de la población bajo estudio. Es decir, más que determinar el punto final de un proceso iterativo hasta que se alcanza la convergencia, puede ser interesante observar la forma en que la población se adapta a su entorno en cada momento, cómo sus individuos se enfrentan mejor al contexto

² Los algoritmos genéticos utilizan procesos de búsqueda o acercamiento al óptimo basados en reglas de transición probabilísticas.

cambiante que les rodea, y cómo, en suma, la población evoluciona; lo que, en última instancia, sugiere un segundo uso alternativo de un algoritmo genético, especialmente útil en el contexto económico: reflejar el modo en que se transforma tal población. En este sentido, los algoritmos genéticos pueden servir como herramienta de predicción de los cambios en la composición de una población³.

Cuando el algoritmo genético se emplea con la finalidad de optimizar o mejorar el resultado de alguna función, la población inicial suele estar formada por cadenas binarias generadas aleatoriamente, a cada una de las cuales corresponde un determinado valor de la función que se desea optimizar. Además, la nueva población suele ser creada sobre la base de las probabilidades de actuación asignadas a determinados operadores. Sin embargo, si se pretende aplicar un algoritmo genético sobre una población económica conocida con el objetivo de predecir su composición en el futuro, la estructura básica del algoritmo genético —en lo que se refiere a la propia transformación de la población inicial en final— debería ser modificada para lograr resultados más apropiados al conocimiento del que se disponga sobre el entorno. Como señala Davis (1991), para derivar un algoritmo genético con un alto grado de eficacia en un contexto determinado es conveniente especializarlo en dicho contexto, lo que conlleva el uso de parámetros y operadores especialmente adaptados al problema en cuestión. Así, en un entorno real, la definición de cadenas binarias puede ser insuficiente para recoger las características relevantes de los individuos de la población. Además, la obtención de la nueva población a partir de la actuación de operadores estocásticos convencionales podría, por una parte, conducir a la aparición de individuos con características incoherentes, y, por otra, impedir la consideración explícita de que

³ Los algoritmos genéticos se han utilizado con el fin de predecir el valor de una magnitud económica (véase, por ejemplo, Mahfoud y Mani, 1996; Álvarez-Díaz y Álvarez, 2002; o Venkatesan y Kumar, 2002). Sin embargo, no se tiene constancia del uso de esta herramienta con la finalidad de

determinadas transformaciones de unos individuos en otros pueden ser más probables que otras. Desde este punto de vista, parece necesario sustituir los operadores genéticos tradicionales por otro mecanismo que sea capaz de incorporar una mayor dosis de información cualitativa sobre el entorno concreto.

El objetivo de este documento es justificar la utilidad de los algoritmos genéticos como herramienta de predicción de la composición de determinadas poblaciones y, en concreto, diseñar un algoritmo genético basado en la actuación de una matriz de transición como fórmula para incorporar explícitamente la información específica del entorno que pueda orientar el proceso de transformación. Para ello, el texto se estructura en las siguientes secciones. En la primera, se presenta la estructura habitual de los algoritmos genéticos y se definen los operadores genéticos en su versión más simple. A continuación, se sugieren modificaciones teóricas que sería conveniente incluir si se desea usar esta técnica con fines predictivos en un contexto particular y se describe la forma y actuación de la matriz de transición en el ámbito de la predicción de una población concreta de agentes económicos. En la sección cuarta se realiza una aplicación de los algoritmos genéticos diseñados para las poblaciones de turistas que visitan las zonas norte y sur de Tenerife en las temporadas de verano e invierno. El documento finaliza con la exposición de las conclusiones obtenidas.

2. ESTRUCTURA CONVENCIONAL DE LOS ALGORITMOS GENÉTICOS

El algoritmo genético es una herramienta capaz de transformar una población original, cuyos individuos están identificados por un vector de características o cadena, en otra población final, formada por un cierto número de individuos —no necesariamente coincidente con el de la población original—, la mayor parte de los

predecir cambios en la composición de una población de agentes económicos, excepto en Hernández-López (2002, 2004a, 2004b).

cuales poseen, previsiblemente, características similares a aquéllos que, en la población original, parecían mejor adaptados al entorno.

Es decir, en el momento temporal t se observa la población inicial, Ω_t , integrada por n individuos cuyas características están representadas a través de cadenas codificadas. Cada cadena es evaluada para tener una medida de su adecuación al contexto a través de la función objetivo que se pretende mejorar, llamada función de calidad o *fitness function*.

La población final en el momento $t+1$, Ω_{t+1} , se forma seleccionando, en primer lugar, los individuos más adaptados de Ω_t , definiendo el grado de adaptación del individuo como el valor de la función de calidad correspondiente a dicho individuo. Después de esta fase, en la que se obtiene una población intermedia cuyos individuos tienden a poseer mayor calidad media que en la población original, se aplican los operadores genéticos para provocar la transformación de la población intermedia en la población final. Estos pasos se reiteran hasta que se cumpla algún criterio de parada⁴:

Es decir, el funcionamiento típico de un algoritmo genético es el siguiente. Una vez que la codificación de las cadenas ha sido establecida, la población inicial ha quedado determinada y la función de calidad ha sido definida, la transformación de una población inicial en otra final se deja en manos del procedimiento de selección y los llamados operadores genéticos, que actúan de acuerdo con reglas probabilísticas.

La propia supervivencia de los mejores, la imitación de la herencia genética que se produce de generación en generación, o la incorporación de nuevo material genético debido a mutaciones en los *bits* de las cadenas, son principios que inspiran los diferentes operadores. La mayoría de ellos fueron planteados originariamente por Holland (1975) y popularizados y modificados por Goldberg (1989). Precisamente, el algoritmo

genético más popular a niveles aplicados, es el algoritmo genético simple de Goldberg (1989), en el que se hace uso, al igual que en el presente documento, el procedimiento de selección, y los operadores de cruce y mutación. Para una visión más detallada del funcionamiento de tales operadores consúltense la referenciada obra de Goldberg (1989).

Si bien han ido apareciendo definiciones sobre nuevos operadores⁵, lo cierto es que la inclusión de éstos en la estructura básica de un algoritmo genético no ha admitido ningún tipo de duda en la literatura. En todos los casos, se trataba de mejorar las propiedades del algoritmo⁶. Sin embargo, y como apuntaba De Jong (1985) en una revisión sobre la historia de la investigación en torno a los algoritmos genéticos, tampoco parecía haberse mejorado sustancialmente las ideas originales de Holland (1975). Por ejemplo, no se había resuelto adecuadamente el problema de asignar valores iniciales a las probabilidades de actuación de los operadores genéticos y la modificación de dichos valores a lo largo de la ejecución de un algoritmo genético.

3. ALGORITMOS GENÉTICOS Y PREDICCIÓN. CARACTERÍSTICAS POBLACIONALES Y MATRIZ DE TRANSICIÓN

La transformación de una población en otra a través de una combinación del procedimiento de selección y los operadores genéticos parece apropiada cuando las poblaciones se generan aleatoriamente, cada individuo está caracterizado por variables binarias, y se aplican con determinadas probabilidades sin suponer que determinados

⁴ Consúltense las obras de Holland (1975), Goldberg (1989), Michalewicz (1994), Dawid (1996), y Moreno y Moreno (1999).

⁵ Además de los operadores genéticos básicos, existen otros que han sido usados en distintas aplicaciones. Entre los más conocidos se encuentra el operador *inversión* (Holland, 1975), el operador *elección* (Arifovic, 1989) y los operadores de *empalme* y *de corte* (Goldberg y otros, 1989).

⁶ Baker (1987) señalaba que un algoritmo genético debe generar poblaciones exactas, consistentes y eficientes: exactas, en el sentido de que no hubiera diferencia —o hubiera la menor posible— entre la población real y la esperada, es decir, el sesgo debería ser bajo; consistentes o precisas, en el sentido de que la propagación —o el rango de valores posibles para el número de descendientes de cada individuo— fuera la menor posible; y eficientes, desde el punto de vista de que no provocaran incrementos en la complejidad del algoritmo genético.

cruces o mutaciones son más verosímiles que otros. Sin embargo, en contextos reales y conocidos, la aplicación mecánica de estas definiciones puede conducir a la generación de poblaciones de individuos cuyas características son incoherentes. Además, puede existir información de tipo cualitativo que sugiera la mayor verosimilitud de unas determinadas transformaciones frente a otras, información que sería interesante incorporar en la ejecución del algoritmo. En casos como éstos, es necesario redefinir las cadenas que identifican a los individuos de la población en términos de caracteres no binarios. Entonces, es imprescindible replantear la definición de operadores de cruce y mutación. Y, sobre todo, quizás resulte apropiado mantener el operador selección para determinar una población intermedia del tamaño deseado. Pero, una vez determinadas las copias, el planteamiento idóneo sería conocer los valores p_{ij} de una hipotética matriz de transición M que, para cada individuo i observado en la población intermedia resultante de las copias, indique la probabilidad de que dicho individuo se transforme en el individuo j , entre todos los posibles que pudieran ser considerados por combinación de los distintos valores de cada una de las características o variables que los definen.

Antes incluso de construir el algoritmo, su diseño y actuación están condicionados por la definición de los individuos de la población. En un contexto real, carece de sentido que la población inicial se genere aleatoriamente, pero, además, las cadenas o estructuras que identifican a los individuos deben definirse apropiadamente. Algunas de las características individuales serán atributos policotómicos difíciles de recoger con cadenas binarias si se pretende evitar inconsistencias lógicas sin una pérdida notable de información.

Obviamente, es necesario establecer una función de calidad que determine el grado de adaptación al entorno de cada individuo en función de sus características. En poblaciones reales, esta función suele ser desconocida de antemano, en cuyo caso tendrá

que ser estimada. Y en dicho proceso de estimación, la conversión de una característica multimodal en una variable dicotómica puede introducir sesgos considerables como consecuencia de la pérdida de información. La solución elemental consiste en definir tantas variables dicotómicas como modalidades posea la característica original. Es decir, si el individuo i está definido por el vector de características $\mathbf{X}_i : (X_{i,1}, \dots, X_{i,k})$ y la característica $X_{i,g}$ posee m modalidades, la información proporcionada por $X_{i,g}$ puede sustituirse por la que aportan m variables dicotómicas $Y_{i,g,1}, \dots, Y_{i,g,m}$ cada una de las cuales toma el valor 1 si la modalidad es una determinada y el valor cero en otro caso. Ahora bien, si la cadena o vector de características que define al individuo se expresa en términos de estos conjuntos de variables dicotómicas, la aplicación de los operadores de cruce y mutación puede conducir a inconsistencias lógicas. En las m posiciones de la cadena que identifica la característica $X_{i,g}$ debe aparecer un 1 en la posición correspondiente a la modalidad que caracteriza al individuo i y el valor cero en el resto de posiciones. Pero esta disposición no está garantizada cuando se aplican los operadores de cruce o mutación.

La solución a problemas de este tipo exige rediseñar el algoritmo de forma que el operador cruce actúe directamente sobre cadenas no binarias. Evidentemente, dado que el carácter $x_{i,g}$ puede transformarse en cualquier otro de los valores del rango de X_g , sería necesario también redefinir el operador mutación de forma que

$$\sum_{\substack{x'_{i,g} \in R_g \\ x'_{i,g} \neq x_{i,g}}} P(m(x_{i,g}) = x'_{i,g}) = \mu .$$

En cualquier caso, y aunque no se produzcan inconsistencias, si se pretende que la dinámica de transformación de la población original se adapte al proceso real, los operadores de mutación y cruce difícilmente pueden discriminar transformaciones de acuerdo con su naturaleza de modo que unas sean más probables que otras. Desde este

punto de vista, y como ya se comentó, la matriz de transición es una herramienta adecuada para este fin. A continuación, se define con precisión el mecanismo de actuación de esta matriz.

Supóngase que el individuo i , $i=1,\dots,n$, puede transformarse en el individuo j , $j=1,\dots,m$, con probabilidad p_{ij} . En general, m es igual a n , pero se ha incluido el parámetro m para recoger el caso en que sólo m de los n individuos se diferencien en, al menos, una característica. Supóngase, entonces, que existen m individuos cuyas cadenas representativas o estructuras, E_1,\dots,E_m , son diferentes⁷. Se asume que cada estructura E_i , con $i=1,\dots,m$, puede transformarse en otra estructura E_j , con $j=1,\dots,m$, con probabilidad p_{ij} . Por supuesto, p_{ii} es la probabilidad de que las características del individuo no cambien. Estas probabilidades pueden recogerse en una matriz cuya fila i -ésima contenga los valores p_{ij} , $j=1,\dots,m$, que indican la probabilidad de que la estructura E_i se convierta en cada una de las m estructuras posibles.

Con la finalidad de ilustrar el proceso de determinación de las probabilidades de transformación, p_{ij} , supóngase que el conocimiento que se tiene de la población real en el momento t no sugiere grandes cambios para una generación posterior $t+1$. En este caso, una forma de establecer las probabilidades de transición sería considerar que p_{ij} es inversamente proporcional al número de *bits* distintos entre las estructuras E_i y E_j . Es decir, si existe un *bit* de diferencia, $p_{ij} = \beta$; $p_{ij} = \beta/2$ si existen dos *bits* diferentes; $p_{ij} = \beta/3$ si las dos estructuras se diferencian en 3 *bits*, etcétera; y si entre dichas estructuras no existe ningún *bit* de diferencia, $p_{ij} = \alpha$, $i,j=1,\dots,m$. En general, se tiene que $p_{ij} = \beta/\delta_{ij}$, en el caso de que las estructuras E_i y E_j difieran en δ_{ij} *bits* o caracteres,

⁷ Nótese que, en una población real, varios individuos diferentes pueden tener las mismas características y, por tanto, estar definidos por la misma cadena o estructura.

cumpléndose que la suma por filas debe ser igual a la unidad, es decir,

$$\sum_{j=1}^m p_{ij} = 1, \quad \forall i = 1, \dots, m^8$$

Teóricamente, la matriz de transición será una matriz cuadrada y simétrica cuyo número de filas y columnas coincide con el número total de estructuras diferentes posibles —según el número de variables y el número de sus correspondientes modalidades— que se podrían observar en la población inicial. Sin embargo, en determinadas aplicaciones reales de gran dimensión, la información disponible sobre la población puede aconsejar la no consideración de algunas estructuras. Además, puede que dicha información también aconseje limitar las estructuras distintas de la población final del momento $t+1$ a aquéllas observadas en el momento t^9 . Manteniendo esta hipótesis, la dimensión de la matriz de transición se reduce considerablemente, y el valor de β no es el mismo de una fila a otra de la matriz, ya que varía en función del número de similitudes entre cada estructura E_i observada en el momento t y las restantes estructuras observadas en dicho momento, únicas candidatas a la transformación, si la hubiere.

Una vez ha quedado establecida la matriz de transición, el algoritmo que hace uso de tal herramienta se ejecuta en dos fases. En la primera, se generan los individuos de la población intermedia resultante de las copias. En la segunda, se generan los individuos de la población final utilizando la matriz de transición. En primer lugar, se construye la matriz de transición, cuya fila i -ésima indica las probabilidades de que la estructura E_i se transforme en cada una de las m estructuras E_j posibles. Luego, se

⁸ Véase Hernández-López (2004b).

⁹ Es preciso admitir que esta restricción impide la aparición de nuevo material genético, pero de este modo se obtienen ventajas computacionales que permitirían la ejecución práctica del algoritmo en ordenadores personales, en casos en los que los individuos hayan sido definidos mediante un importante número de características y cada una de éstas pueda presentar un número elevado de modalidades. Obviamente, esta suposición sólo podría ser admitida si la información del contexto así lo aconsejara. No se pretende hacer de la misma una hipótesis generalizable a cualquier caso.

identifica cada uno de los n individuos, $I_{2,q}$, con $q=1,\dots,n$, de la población intermedia con alguna de las estructuras E_i , $i=1,\dots,m$. Finalmente, para determinar en qué nuevo individuo $I_{3,q}$, $q=1,\dots,n$, se transforma cada uno de los individuos $I_{2,q}$ —identificado con alguna estructura E_i —, se genera una prueba multinomial cuyos resultados son las estructuras E_j en que puede transformarse $I_{2,q}$ y cuyas probabilidades son las de la fila i -ésima de la matriz de transición correspondiente.

Se podría pensar que un adecuado control de las probabilidades de actuación de los operadores genéticos evitaría la introducción de la matriz de transición. No obstante, la matriz de transición permite una mayor flexibilidad al posibilitar probabilidades distintas para cada transformación de las cadenas iniciales, algo que no ocurre con la utilización de los operadores genéticos, cuyas probabilidades de actuación son aplicables por igual a cualquier individuo de la población sin atender a la mayor o menor versomilitud de ocurrencia de cada una de las transformaciones de las estructuras integrantes de la población inicial.

El problema de esta nueva aproximación radica en que la dimensión de la matriz puede ser muy grande, y además, en que no existe un criterio único para determinar los valores p_{ij} , aunque la incorporación de información cualitativa y del conocimiento más profundo de la población puede hacer que estos problemas no sean, finalmente, tan graves.

Supóngase, por ejemplo, que se trata de predecir la composición de la población que visita un determinado destino turístico¹⁰. Cabe suponer que dicha población experimenta cambios de una temporada a la siguiente, de modo que los turistas más satisfechos probablemente repetirán su visita, mientras que aquéllos no satisfechos difícilmente volverán. En este sentido, el grado de satisfacción puede cumplir la labor

de la función de calidad y ese grado dependerá de características del turista tales como la nacionalidad, la renta, la edad, el gasto realizado o el régimen de alojamiento. Probablemente, en la temporada siguiente aumentará la participación de los individuos con características similares a aquéllos que en la temporada anterior mostraron un grado de satisfacción más elevado. Este principio puede guiar la aplicación del procedimiento de selección. Pero, por otra parte, parece lógico pensar que es más probable que un turista alemán, con alta renta, de edad mediana y alojado en hotel de 5 estrellas sea “sustituido” por un inglés de alta renta, edad mediana y alojado en hotel de 5 estrellas que por un francés de baja renta, joven y alojado en apartamento.

Este tipo de informaciones cualitativas pueden incorporarse explícitamente en la matriz de transición y, en cambio, no se tomarían en consideración si se emplean los operadores convencionales de mutación y cruce. De este modo, el algoritmo genético permite predecir el modo en que va cambiando la población de turistas de acuerdo con características relevantes para los gestores públicos y privados de esa rama de actividad en el área geográfica correspondiente.

4. APLICACIÓN DE LOS ALGORITMOS GENÉTICOS PARA LA PREDICCIÓN DE LA COMPOSICIÓN DE LA POBLACIÓN DE TURISTAS EN TENERIFE

En este epígrafe se aplican dos tipos de algoritmos genéticos, uno que incorpora los operadores convencionales y otro que introduce la matriz de transición, para predecir los cambios en la composición de las poblaciones de turistas que visitan el norte o el sur de Tenerife en verano o en invierno. Para ello, se describen, en primer lugar, las características de estas poblaciones así como las resultantes de las ejecuciones de los algoritmos diseñados, en términos de la composición y calidades medias en las

¹⁰ En un entorno turístico, Hurley *et al.* (1998) proponen un algoritmo genético cuya aplicación se refiere al problema de la localización de almacenes de venta al por menor. Sin embargo, el algoritmo no

poblaciones predichas en comparación con las propias de las poblaciones originales. En segundo lugar, se evalúa la capacidad predictiva del algoritmo, a través de un simple test de bondad de ajuste entre la población predicha por el algoritmo y la población realmente observada en el año siguiente a aquél para el que se tomaron los datos de la población original. En estos términos, se compara la capacidad predictiva del algoritmo genético simple frente a la del algoritmo genético con matriz de transición.

Con estos objetivos y gracias al expreso consentimiento del Excelentísimo Cabildo Insular de Tenerife, que autorizó el acceso a un grupo de encuestas a turistas seleccionadas entre las realizadas en la temporada 1997/1998, se obtuvo la base de datos que constituye la materia prima de la presente investigación. En concreto, se consideró necesario dividir la población objeto de estudio según zona de alojamiento — distinguiendo la denominada zona sur, integrada por el espacio costero desde Radazul hasta Los Gigantes, y el resto de la Isla—, así como atendiendo a la temporada en que se produce la visita. Es decir, los turistas encuestados en la temporada 1997/1998 se dividieron en 4 subpoblaciones: dos temporadas —invierno y verano¹¹— y dos zonas —“costa sur” y “resto”. Y dentro de cada una de estas subpoblaciones el método de muestreo aplicado para la petición de las encuestas fue de tipo estratificado por nacionalidad y mes.

Cada turista encuestado respondió a cuestiones referentes a edad, residencia, tipo de alojamiento, categoría del alojamiento, servicios contratados, etc., que permiten identificar el tipo de turista y, *a priori*, pueden considerarse variables explicativas del grado de satisfacción que éste obtiene de su visita. Tal grado de satisfacción puede deducirse a partir de otras cuestiones que tratan de reflejar la percepción del turista en relación a aspectos tales como el paisaje, el sol, las playas, el trato recibido, etc..

se emplea con la finalidad de predecir cambios en la composición de la población de turistas.

El resumen de estos *ítems* en un único indicador general del grado de satisfacción del turista i -ésimo consistió en una media ponderada, es decir:

$$F_i = \frac{\sum_{m=1}^j w_m F_{i,m}}{\sum_{m=1}^j w_m}, \quad m = 1, \dots, j; \quad i = 1, \dots, N$$

donde $F_{i,m}$ es la variable que mide el grado de satisfacción del turista i ($i=1, \dots, N$) en relación con el aspecto m ($m=1, \dots, j$); y w_m es la ponderación con que la variable $F_{i,m}$ entra a formar parte del promedio. Se ha considerado adecuado que estas ponderaciones sean las puntuaciones de los j *ítems*¹² en el primer factor resultante de la aplicación de un análisis de componentes principales.

Entonces, y asumiendo que ese grado de satisfacción depende de un conjunto de características individuales, se procedió a estimar los parámetros de una función lineal de tales características que diera como resultado un grado de satisfacción recibido por el turista.

Bajo el supuesto de que el grado de satisfacción percibido por el turista i -ésimo, F_i , depende de un vector de k variables explicativas, $\mathbf{X}_i: \{X_{i,1}, \dots, X_{i,k}\}$, que recogen características específicas de dicho turista, el siguiente paso para la aplicación del algoritmo consiste, entonces, en formular una relación funcional —*fitness function*— que explique el valor de F_i a partir de las variables X_i , de tal manera que pueda predecirse el cambio de calidad de la población cuando se modifica la estructura de la misma.

Dado que en las cuestiones relativas a la valoración de la calidad se establece una puntuación mínima de 0 y máxima de 10, se asumió que la variable aleatoria

¹¹ Por temporada de invierno se considera el período comprendido entre los meses de octubre de un año y marzo del siguiente año. La temporada de verano abarca desde abril a septiembre de un mismo año.

¹² En general, los *ítems* sobre los que el turista tuvo que expresar su grado de satisfacción fueron los siguientes: paisaje, playas, sol, temperatura, clima, calidad medioambiental, estética del centro de vacaciones, calidad del alojamiento, precios, trato recibido en la Isla, seguridad, tranquilidad, alimentación, compras, piscinas, diversión, información turística y trato recibido en el alojamiento.

$F_i / \mathbf{X}_i = \mathbf{x}_i$ sigue una distribución normal truncada en el intervalo [0,10], cuyo valor esperado viene dado por:

$$E\left[\frac{F_i}{\mathbf{X}_i = \mathbf{x}_i}\right] = \beta' \mathbf{x}_i + \sigma \frac{\phi\left(\frac{-\beta' \mathbf{x}_i}{\sigma}\right) - \phi\left(\frac{10 - \beta' \mathbf{x}_i}{\sigma}\right)}{\Phi\left(\frac{10 - \beta' \mathbf{x}_i}{\sigma}\right) - \Phi\left(\frac{-\beta' \mathbf{x}_i}{\sigma}\right)}$$

donde ϕ y Φ son, respectivamente, las funciones de densidad y distribución de la normal estándar y $\beta' \mathbf{x}_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$.

Aunque el documento presente es fruto de un estudio que analiza las 4 subpoblaciones consideradas, por evidentes razones de espacio, en este documento as páginas sólo se ha considerado oportuno presentar únicamente los resultados para una de ellas, concretamente, la que hace referencia a los turistas alojados en la zona sur de la Isla durante los meses de verano de 1998¹³.

Las características que definen a los individuos de la población se recogieron a través de variables cualitativas que toman el valor 1 en una situación específica y el valor 0 en el resto de casos. Dichas variables se definieron a partir de las variables originales de la encuesta en función de la significación estadística de las diferencias de calidad observadas entre las situaciones recogidas por unas y otras. Pero una vez que quedó determinada la función de calidad —véanse definiciones de variables binarias y estimaciones de la función de calidad en el cuadro 1—, las variables significativas en la misma fueron las que se utilizaron para identificar a los individuos pertenecientes a la subdivisión temporal y zonal considerada.

¹³ Para un análisis de los resultados correspondientes a las otras 3 subpoblaciones consideradas, consúltese Hernández López (2002).

CUADRO 1. *Definición de regresores (variables binarias) y estimación de parámetros de la función de calidad.*

Nombre	Definición	Estimación
X2A	Procedencia del Reino Unido, Alemania o Francia	1.63950
X2B	Procedencia de Bélgica	1.68635
X2C	Procedencia de España, Italia o Suiza	-
X3A	Número de pernoctaciones inferior o igual a 7 días	-1.05357
X3B	Número de pernoctaciones superior a 7 días e inferior que 14	-
X3C	Número de pernoctaciones superior o igual a 14 días	-
X5A	Alojamiento en hotel de 3 o 4 estrellas, <i>aparthotel</i> de 2 o 4 estrellas o <i>time-sharing</i>	-0.90636
X5B	Alojamiento en <i>aparthotel</i> de 3 estrellas o apartamento de 3 llaves	-2.63475
X5C	Alojamiento en apartamento de 1 o 2 llaves, casa o apartamento privado	-
X6A	Viaje motivado por recomendación de amigos, visitas anteriores o no existió motivo específico	-0.85742
X6B	Viaje motivado por recomendación de agencia de viajes	-
X6C	Viaje motivado por la información contenida en un folleto de agencia de viajes	-
Constante		8.69252
Desviación típica		0.76063

Sin embargo, la ejecución del algoritmo sobre las cadenas que identificaban a los turistas que integraban cada una de las poblaciones bajo estudio requería la recodificación en formato no binario de las variables —consúltese el cuadro 2 a este respecto.

CUADRO 2. *Recodificación no binaria de las variables que identifican a los turistas.*

Nombre	Modalidades
Y1	1 (Reino Unido, Alemania o Francia); 2 (Bélgica); 3 (España, Italia o Suiza)
Y2	1 (pernoctaciones ≤ 7 días); 2 (pernoctaciones $\in (7,14)$); 3 (pernoctaciones ≥ 14)
Y3	1 (hotel de 3 o 4 estrellas, <i>aparthotel</i> de 2 o 4 estrellas o <i>time-sharing</i>); 2 (<i>aparthotel</i> de 3 estrellas o apartamento de 3 llaves); 3 (apartamento de 1 o 2 llaves o casa o apartamento privado)
Y4	1 (recomendación de amigos, por visitas anteriores o no existió motivo específico); 2 (recomendación de agencia de viajes); 3 (folleto de agencia de viajes)

4.1 Descripción de la población original

En los cuadros 3 y 4 se indican las características básicas de la población en cuestión. Mayoritariamente, este grupo de turistas estaba compuesto por británicos, alemanes o franceses; cuya estancia no superaba la semana; se trataba, en general, de turistas alojados en hoteles de 3 o 4 estrellas, *aparthoteles* de 2 o 4 estrellas, o en casas

o apartamentos en régimen de propiedad compartida que llegaron impulsados por la recomendación de amigos, sus propias visitas anteriores o sin motivo especial. En términos medios, el grado de satisfacción es algo menor que en las poblaciones precedentes —con un valor de 7.5527—, lo que también se aprecia, obviamente, si se observa el valor de la calidad media alcanzada por los grupos de turistas definidos mediante la combinación del país de procedencia con las modalidades de las otras características, como muestra el cuadro 4.

CUADRO 3. Población original. Información porcentual y de calidades medias por modalidad de variable

		Frecuencia (%)	Calidad media	
			Observada	Ajustada
Y2	1	76.6	7.911848	7.911847
	2	2.1	8.555340	8.555340
	3	21.3	6.159806	6.159804
Y3	1	42.6	7.013463	7.013464
	2	19.1	7.470478	7.612603
	3	38.3	8.193132	8.122065
Y5	1	59.6	7.902681	7.902680
	2	27.7	6.355390	6.355388
	3	12.8	8.514132	8.514133
Y6	1	78.7	7.620747	7.620746
	2	4.3	7.506796	7.168925
	3	17.0	7.249838	7.334304

CUADRO 4. Población original. Calidad media observada según país de procedencia y resto de variables

Calidad media poblacional: 7.5527		Y2		
		1 R.U., Alem., Franc.	2 Bélgica	3 Esp., Ital., Suiz.
Y3	1	7.3173 (7.3525)	-	5.7979 (5.6574)
	2	8.2388 (8.1801)	8.5553 (8.5553)	5.8283 (6.3525)
	3	8.4370 (8.4191)	-	6.9739 (6.6369)
Y5	1	8.3115 (8.3263)	-	6.6413 (6.5990)
	2	6.7511 (6.7215)	8.5553 (8.5553)	5.0362 (5.1349)
	3	8.5141 (8.5141)	-	-
Y6	1	8.0193 (8.0214)	8.5553 (8.5553)	6.1091 (6.1015)
	2	7.5068 (7.1689)	-	-
	3	7.5455 (7.6481)	-	6.3628 (6.3930)

Nota: Entre paréntesis se indica la calidad media estimada. Las casillas en las que no se indica valor alguno corresponden a situaciones en las que no se ubicó ninguno de los turistas.

Teniendo en cuenta dichas calidades, se puede indicar que un grado de satisfacción medio estimado inferior a 6 correspondió a los turistas españoles, italianos o suizos con estancias más cortas o que se alojaron en *aparthotel* de 3 estrellas o apartamento de 3 llaves. Por el contrario, los grupos compuestos por turistas belgas — con estancias entre una y dos semanas, alojados en *aparthotel* de 3 estrellas o apartamento de 3 llaves o que eligieron la Isla como lugar de vacaciones por recomendación de amigos, visitas anteriores o sin motivo específico— fueron los más satisfechos.

4.2. Resultados de las ejecuciones obtenidas

Establecido el marco de referencia para comprobar los cambios predichos por los algoritmos genéticos en las poblaciones descritas, se exponen los resultados de las ejecuciones, en primer lugar, de un algoritmo genético simple diseñado con los tradicionales operadores de selección, cruce y mutación, y, en segundo lugar, del algoritmo genético que ha sido diseñado e implementado con el objetivo de que se aplique una matriz de transición a la población intermedia resultante de las copias efectuadas de acuerdo con el operador selección.

Las características univariantes de la población generada por el algoritmo genético para la temporada 1998/1999 correspondientes a la subpoblación considerada a través del presente documento, quedan recogidas en el cuadro 5. En este caso, los resultados presentados se refieren a la ejecución del algoritmo con operador selección tipo proporcional, probabilidad de cruce igual a 0.6 y probabilidad de mutación igual a 0.001.¹⁴

¹⁴ Se ha seleccionado la población que, *a posteriori*, ha resultado ajustarse mejor a la población realmente observada en la campaña 1998/1999. El grado de desajuste se ha medido a través del estadístico de contraste de bondad de ajuste entre las dos poblaciones en cuestión.

CUADRO 5. Composición de la población y calidad media estimada por modalidad de variable en la población predicha por el algoritmo genético simple (selección proporcional, probabilidad de cruce 0.6, probabilidad de mutación 0.001)

		Frecuencia (%)	Calidad media estimada
Y2	1	89 (76.6)	8.4486 (7.911847)
	2	3 (2.1)	8.4989 (8.555340)
	3	8 (21.3)	7.4490 (6.159804)
Y3	1	58 (42.6)	8.3188 (7.013464)
	2	13 (19.1)	8.5959 (7.612603)
	3	29 (38.3)	8.3716 (8.122065)
Y5	1	42 (59.6)	8.3413 (7.902680)
	2	0 (27.7)	- (6.355388)
	3	58 (12.8)	8.3911 (8.514133)
Y6	1	90 (78.7)	8.3594 (7.620746)
	2	0 (4.3)	- (7.168925)
	3	10 (17.0)	8.4666 (7.334304)

Nota. Entre paréntesis se indican los valores correspondientes a la población original.

Algunas transformaciones observadas se ajustan a lo esperado. Por ejemplo, desciende notablemente la participación de los turistas procedentes de España, Italia y Suiza; y aquéllos que se alojaban en *aparthotel* de 3 estrellas o apartamento de 3 llaves, y que representaban más de la cuarta parte de la población original, ni siquiera están presentes en la población final.

También cabe señalar el crecimiento de la participación de los turistas que se decantaron por este destino turístico debido a la recomendación de amigos, por visitas anteriores o sin que existiera ningún motivo específico.

En cuanto al número de pernотaciones, se vuelve a registrar un incremento en la dirección contraria a la esperada, puesto que aumenta la participación de las estancias cortas a costa de las estancias más largas, a pesar de que los turistas que permanecieron en la Isla menos de 7 días fueron los menos satisfechos en la población original y seguirían siéndolo en la población final generada por el algoritmo.

CUADRO 6. *Calidad media estimada según país de procedencia y resto de variables en la población predicha por el algoritmo genético simple (selección proporcional, probabilidad de cruce 0.6, probabilidad de mutación 0.001)*

Calidad media poblacional: 8.3702		Y2		
		1 R.U.,Alem.,Franc.	2 Bélgica	3 Esp.,Ital.,Suiz.
Y2	1	8.3445 (7.3525)	8.4167 (-)	6.7815 (5.6574)
	2	8.6061 (8.1801)	8.5400 (8.5553)	- (6.3525)
	3	8.6349 (8.4191)	-	7.4490 (6.6369)
Y3	1	8.4786 (8.3263)	8.5400 (8.5553)	7.4970 (6.5990)
	2	- (6.7215)	-	- (5.1350)
	3	8.4301 (8.5141)	-	7.3050 (-)
Y4	1	8.4143 (8.0214)	8.4989 (8.5553)	7.1168 (6.1015)
	2	- (7.1689)	-	-
	3	8.9236 (7.6481)	-	7.7812 (6.3930)

Nota: Entre paréntesis se indica la calidad media estimada en la población original.

Combinando el país de procedencia con las modalidades del resto de variables —véase cuadro 6—, es posible apreciar la desaparición de varios grupos de turistas con baja calidad en la población original, como ocurre, por ejemplo, con los españoles, italianos y suizos que se alojaron en *aparthotel* de 3 estrellas o en apartamento de 3 llaves, o que disfrutaron de estancias entre una y dos semanas. También aparece un grupo nuevo con elevada calidad compuesto por turistas belgas que pernoctan en la Isla un máximo de 7 días.

Estas transformaciones en la distribución del número de pernoctaciones según país de procedencia pueden estar relacionadas con el movimiento, antes comentado, experimentado por las modalidades relativas a la duración de la estancia. Finalmente, llama la atención la desaparición del grupo de turistas belgas que se alojaron en *aparthotel* de 3 estrellas o apartamento de 3 llaves.

Aunque los resultados no se muestren aquí por las comentadas cuestiones de espacio, la combinación de diferentes tipos de operador selección y distintas probabilidades de actuación de los operadores de cruce y mutación sobre la misma población han producido notables diferencias en la composición de la población final.

Esta sensibilidad de los resultados a la elección de los parámetros que gobiernan el funcionamiento del algoritmo, invita a adoptar algún criterio que, sin dejar de ser operativo, reduzca la arbitrariedad del efecto distorsionador que se permite a los operadores de cruce y mutación.

Este es el objetivo que se pretende alcanzar con la introducción de información económica propia del contexto en el que se han aplicado los algoritmos genéticos propuestos. En este sentido, se consideró oportuno incorporar una matriz de transición que dirija la transformación final de los individuos que componen la población intermedia resultante de la aplicación del operador selección, sustituyendo la acción de los operadores genéticos en la estructura habitual de un algoritmo genético. La información cualitativa que permite incorporar tal matriz se introduce en el algoritmo a través de las probabilidades de transformación de cada uno de los tipos de turista observados en la temporada 1997/1998 en cada uno de los tipos de turista que se consideran admisibles.

Como los operadores de cruce y mutación han sido sustituidos por las probabilidades de transformación de cada estructura observada recogidas en la matriz, el margen de maniobra en la ejecución se limita a la elección del tipo de operador selección: proporcional o elitista. Eligiendo, de nuevo, en cada temporada y zona de alojamiento, aquella población que, *a posteriori*, se ajustase mejor a la población realmente observada en la campaña 1998/1999, los comentarios que siguen se refieren a las características de estas poblaciones obtenidas mediante el algoritmo genético que incorpora la matriz de transición.

En el caso de la subpoblación alojada en la zona sur durante la temporada de verano, la ejecución elegida, en este caso, ha sido la que incorpora operador selección de tipo proporcional —véanse cuadros 7 y 8.

CUADRO 7. *Composición de la población y calidad media estimada por modalidad de variable en la población predicha por el algoritmo genético que incorpora una matriz de transición (selección proporcional)*

		Porcentaje	Calidad media estimada
Y2	1	87 (76.6)	8.2751 (7.9118)
	2	3 (2.1)	8.5553 (8.5553)
	3	10 (21.3)	6.3505 (6.1598)
Y3	1	59 (42.6)	8.0366 (7.0135)
	2	12 (19.1)	8.3930 (7.6126)
	3	29 (38.3)	8.0769 (8.1221)
Y5	1	41 (59.6)	8.1698 (7.9027)
	2	10 (27.7)	6.2493 (6.3554)
	3	49 (12.8)	8.4011 (8.5141)
Y6	1	91 (78.7)	8.1494 (7.6207)
	2	1 (4.3)	7.6942 (7.1689)
	3	8 (17.0)	7.4766 (7.3343)

Nota. Entre paréntesis se indican los valores correspondientes a la población original.

Algunas de las transformaciones que cabía prever en esta población son el descenso significativo del porcentaje de turistas procedentes de España, Italia o Suiza, el notable incremento de la participación de los turistas alojados en apartamentos de 1 o 2 llaves o en casa o apartamento privado y el crecimiento más moderado del peso de aquéllos que se decantaron por la Isla siguiendo el consejo de los amigos, guiados por su visitas anteriores o sin que existiera ningún motivo específico.

Más difícil resulta, en cambio, anticipar el cambio operado en las modalidades relativas al número de pernoctaciones. Los turistas con estancias más cortas, los menos satisfechos, ganan peso a costa de los turistas que disfrutaban de estancias más largas y que, según lo declarado en la encuesta, estaban más satisfechos.

En el cuadro 8 se observa, además, que el algoritmo genético predice la desaparición de los turistas belgas que se alojasen en *aparthotel* de 3 estrellas o en apartamento de 3 llaves, a pesar de la elevada calidad media del grupo en cuestión; hecho que quizás resulta favorecido por la escasa presencia de individuos de la población original en dicho grupo.

CUADRO 8. *Calidad media estimada según país de procedencia y resto de variables en la población predicha por el algoritmo genético que incorpora una matriz de transición (selección proporcional)*

Calidad media poblacional: 8.0911		Y2		
		1 R.U.,Alem.,Franc.	2 Bélgica	3 Esp.,Ital.,Suiz.
Y3	1	8.1680 (7.3525)	-	5.5849 (5.6574)
	2	8.5152 (8.1801)	8.5553 (8.5553)	6.9286 (6.3525)
	3	8.4525 (8.4191)	-	6.6369 (6.6369)
Y5	1	8.4257 (8.3263)	8.5553 (8.5553)	6.8714 (6.5990)
	2	6.7269 (6.7215)	-	5.1349 (5.1349)
	3	8.4011 (8.5141)	-	-
Y6	1	8.3086 (8.0214)	8.5553 (8.5553)	6.1338 (6.1015)
	2	7.6942 (7.1689)	-	-
	3	7.8490 (7.6481)	-	6.8559 (6.3930)

Nota: Entre paréntesis se indica la calidad media estimada en la población original.

4.3. Evaluación de la capacidad predictiva de los algoritmos genéticos

Una vez que han sido analizadas las características de la población original, así como de las resultantes de las ejecuciones seleccionadas tras la aplicación del algoritmo genético descrito, es preciso mostrar la capacidad predictiva del mismo. Tal evaluación está supeditada a la veracidad de los supuestos que fue necesario realizar para llevar a cabo la predicción, y que se comentan a continuación. Por otra parte, los errores de predicción no deben atribuirse exclusivamente al erróneo funcionamiento predictivo del algoritmo, sino que también pueden haber sido causados por una incorrecta asignación de valores a las probabilidades de transformación de estructuras. Sin olvidar estas consideraciones, también es preciso definir algún criterio que permita evaluar el grado de coincidencia entre la población predicha por el algoritmo y aquella realmente observada en el período de tiempo para el que se realiza la predicción. En este sentido, la capacidad predictiva del algoritmo se ha evaluado mediante un sencillo test de bondad de ajuste que permita calibrar simultáneamente las diferencias entre la participación observada de cada uno de los grupos de turistas considerados en la

subpoblación dada de la campaña 1998/1999 y las participaciones predichas por el algoritmo genético para tal subpoblación.

A la hora de evaluar la capacidad predictiva de los dos algoritmos desarrollados, es preciso admitir la existencia de una notable limitación. Los datos disponibles recogidos por el Excelentísimo Cabildo Insular de Tenerife en su Estadística de Turismo Receptivo en la campaña 1998/1999 —testigo de la evaluación de los resultados— sólo ofrecen información, por temporada y zona de alojamiento, del país de residencia del encuestado y el carácter —hotelero o extrahotelero— del alojamiento, lo que obliga a que la capacidad predictiva de los algoritmos sea evaluada exclusivamente en función de la participación de las modalidades correspondientes a estos dos atributos en la composición de la población final. Además, la categorías de alojamiento y los países de procedencia para los que la Estadística de Turismo Receptivo proporcionaba información no coinciden con las modalidades de estos dos atributos consideradas en la población generada por el algoritmo. Debe tenerse en cuenta que el procedimiento de estimación de la función de calidad condujo a fusionar categorías de establecimientos alojativos heterogéneos y tomar grupos de países como una única procedencia, debido a la homogeneidad de estas modalidades en cuanto a la calidad percibida por los turistas.

Dada, entonces, la imposibilidad de la comparación directa entre los datos predichos por el algoritmo y los realmente observados, fue necesario introducir ciertos supuestos. En particular, se asume que en cada grupo de turistas generado por el algoritmo —e identificados por una combinación dada de tipo de alojamiento y países de procedencia—, se mantiene su composición interna —procedente de la población original— en términos de las proporciones de cada una de las categorías fusionadas. Es decir, cada uno de los grupos generados por el algoritmo se desagrega, conforme a las proporciones observadas en la población original, en categorías que, *a posteriori*,

permiten la comparación con la población observada. Así, en el caso, por ejemplo, de la subpoblación que ha sido testigo a lo largo del documento de trabajo de la ejecución de los algoritmos, las categorías que posibilitaron la evaluación de la capacidad predictiva de los mismos fueron, según la categoría de alojamiento: Hotelero, Extrahotelero, y según el país o grupo de países de procedencia: Reino Unido/Alemania/Francia (grupo 1), Bélgica (grupo 2), España/ Italia/Suiza (grupo 3).

Al margen de los obstáculos anteriores, la validez de la comparación realizada descansa en otro supuesto: la estabilidad de la función de calidad con el transcurso del tiempo. Dicha función, que sirve de termómetro indicativo del grado de satisfacción de los turistas, procede de la muestra de turistas encuestados y seleccionados en cada una de las subdivisiones poblacionales. El grado de satisfacción expresa la relación entre lo esperado por el turista y lo ofrecido por la oferta turística existente, en el momento en el que visitó la Isla, en la zona de alojamiento elegida. La hipótesis admitida consiste en asumir que dicha relación no ha variado lo suficiente como para que deba ser redefinida la función de calidad. Parece razonable mantener este supuesto para predecir la composición de la demanda turística del año siguiente, entre otras razones porque cabe pensar que la oferta no habrá variado sustancialmente y, por tanto, el mismo vector de características definitorias del turista puede significar un valor similar de la calidad percibida. Sin embargo, es evidente que a mayor número de períodos temporales para los que se solicite una predicción, mayor probabilidad de que sea necesario algún cambio en la definición de las variables y los coeficientes de la función de calidad.

Los resultados de los contrastes de bondad de ajuste aplicados sobre las 6 ejecuciones correspondientes a la subpoblación integrada por los turistas alojados en el sur de la Isla durante la temporada de verano han quedado recogidos en el cuadro 9.

CUADRO 9. Resultados del test de bondad de ajuste de las ejecuciones generadas con el algoritmo genético simple

Tipo de selección	Probabilidad de cruce	Probabilidad de mutación	Valor estadístico	$\chi^2_{0.95,3}=7.81$
Proporcional	0	0	23.280	H_A
Elitismo	0	0	16.340	H_A
Proporcional	1	0.001	22.550	H_A
Elitismo	1	0.001	18.510	H_A
Proporcional	0.6	0.001	11.080	H_A
Elitismo	0.6	0.001	15.600	H_A

A la vista de los resultados presentados en este cuadro, puede considerarse que la capacidad predictiva del algoritmo para esta subpoblación queda más en entredicho, aunque no es posible atribuir directamente estos errores al mal funcionamiento del algoritmo. Además de las limitaciones inherentes a la comparación predictiva, es posible que las probabilidades de actuación de los operadores no sean las adecuadas o que la aplicación aleatoria de unas determinadas probabilidades para los operadores cruce y mutación no sea aplicable en este caso, elemento que justificaría, en última instancia, la introducción de la matriz de transición como sustituta de la labor de los operadores genéticos tradicionales. En cualquier caso, los mejores resultados corresponden a la ejecución que contiene la selección proporcional, probabilidad de cruce igual a 0.6, y probabilidad de mutación de 0.001 —véase cuadro 10—, que será la finalmente utilizada en el ejercicio de predicción que más adelante se plantea.

CUADRO 10. Comparación de la población predicha por el algoritmo genético simple (selección proporcional, probabilidad de cruce igual a 0.6, probabilidad de mutación igual a 0.001) con la población realmente observada

País	% Predichos		% Observados		País	% Predichos		% Observados	
	H	EH	H	EH		H	EH	H	EH
G1	35.15	36.85	26.80	38.60	G1	35.15	36.85	26.80	38.60
G2	2.00	0.00	2.80	1.40	G2*	8.00	0.00	12.30	6.90
G3	6.00	0.00	9.50	5.50					

* Estos valores corresponden a las fusiones de grupos 2 y 3, debido a la aparición de celdas cuya frecuencia observada era inferior o igual a 5.

En segundo lugar, se presentan los resultados obtenidos tras la ejecución del algoritmo genético que incorpora la matriz de transición. En este caso, para cada subpoblación, sólo se han barajado 2 posibilidades, según el tipo de operador selección utilizado.

CUADRO 11. *Resultados del test de bondad de ajuste de las ejecuciones generadas con el algoritmo genético que incorpora una matriz de transición*

Tipo de selección	Valor estadístico	Resultados del contraste	
Proporcional	6.9010	$\chi^2_{0.95,3}=7.81$	H₀
Elitismo	8.5870	$\chi^2_{0.95,3}=7.81$	H _A

Una revisión de los resultados permite advertir que el operador selección de tipo proporcional conduce a una mejor predicción, aunque no a mucha distancia, desde el punto de vista del valor del estadístico, del operador selección de tipo elitista. Además, en el caso del operador de tipo elitista, la hipótesis nula de igualdad entre frecuencias predichas y observadas no ha podido ser rechazada al 97.5% de confianza —véanse los resultados de la comparación entre poblaciones predichas y reales en el cuadro 11. Además, los valores del estadístico cuando se hace uso de la matriz de transición son marcadamente mejores que los resultantes con el algoritmo genético simple, algo que no es exclusivo de esta subpoblación, sino que puede ser generalizado al resto de subpoblaciones analizadas. Esta circunstancia parece indicar que las probabilidades de transformación de estructuras incorporadas en la matriz de transición resultan más adecuadas para explicar las transformaciones realmente operadas en la población que la actuación puramente aleatoria de los operadores de cruce y mutación.

4. CONCLUSIONES

Este documento analiza la conveniencia de utilizar los algoritmos genéticos como herramienta predictiva de la composición de poblaciones reales y conocidas. En el contexto económico, la predicción de determinadas poblaciones constituye un objetivo básico para la planificación económica. Pero más relevante aún puede ser la predicción de los cambios en la composición de dicha población en términos de la presencia de individuos con ciertas características. De este modo, la predicción del agregado que proporcionan las técnicas estadístico-econométricas tradicionales queda complementada por la predicción más rica que aportan los algoritmos genéticos.

Sin embargo, con la finalidad de la predicción de poblaciones reales se hace imprescindible algún tipo de modificación en la estructura convencional de un algoritmo genético. La definición de poblaciones cuyos individuos están identificados por características no binarias puede provocar la inconsistencia de los operadores convencionales. Además, en el proceso de transformación de la población original sería interesante incorporar información cualitativa sobre el entorno al que pertenece la población analizada.

En este documento, se ha mostrado que un mecanismo adecuado para este fin consiste en usar una matriz de transición que, una vez que se ha procedido a la selección y se ha configurado el fondo de emparejamientos, sea aplicada sobre cada una de las cadenas integrantes de la población intermedia teniendo en cuenta la probabilidad de que cada una de aquéllas se transforme en cualquiera de las otras posibles. Dado que el proceso de búsqueda está iniciado por el procedimiento de selección, los individuos de la población intermedia tienden a ser los más adaptados al entorno, como se espera en la realidad. Además, dado que la matriz de transición toma en consideración qué

transformaciones son más probables, cabe pensar que la población predicha de esta forma se ajustará mejor a la población real.

Finalmente, la conclusión más evidente que se puede extraer de los resultados obtenidos a partir de los contrastes de bondad de ajuste realizados entre poblaciones realmente observadas y poblaciones predichas por el algoritmo, es que no sólo los algoritmos genéticos constituyen una herramienta capaz de ofrecer un tipo de predicción sobre la demanda turística —su composición interna— que no ofrecen otras técnicas, sino que, además, queda claro que la bondad de estas predicciones es sensible a la elección de los parámetros que determinan las transformaciones de la población bajo estudio y, en este sentido, la utilización de la matriz de transición significa una reducción de la aleatoriedad de tales transformaciones que se ha traducido en una mejora predictiva. En tal sentido, no es demasiado aventurado pensar que la incorporación de información cualitativa, que ayude a decidir los valores de los parámetros que configuran la matriz de transición, contribuirá a mejorar el funcionamiento del algoritmo con finalidad predictiva.

5. REFERENCIAS

ÁLVAREZ-DÍAZ, M. Y A. ÁLVAREZ (2002) «Predicción no-lineal de tipos de cambio. Aplicación de un algoritmo genético», *Documento de Trabajo*, 0205. Departamento de Economía Aplicada. Universidade de Vigo.

ARIFOVIC, J. (1989) «Learning by Genetic Algorithms in Economic Environments», *Working Paper*, 90-001. Santa Fe Institute.

BAKER, J.E. (1987) «Reducing bias and inefficiency in the selection algorithm», *Genetic algorithms and their applications: Proceedings of the Second International*

Conference on Genetic Algorithms. John J. Grenfenstett (Ed.). Lawrence Erlbaum Associates: 14-21. Hillsdale, NJ.

DAVIS, L. (1991) (Ed.) *Handbook of Genetic Algorithm*. Van Nostrand Reinhold. New York.

DAWID, H. (1996) *Adaptive Learning by Genetic Algorithms. Analytical Results and Applications to Economic Models*. Lecture Notes in Economics and Mathematical Systems, 441. Springer.

DE JONG, K. (1975) *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. Doctoral Dissertation. Dissertation Abstracts International. 36(10) 5140B, University Microfilms, n° 76-9381. University of Michigan.

DE JONG, K. (1985) «Genetic Algorithms: A 10 year perspective», *Proceedings of the First International Conference on Genetic Algorithms and their Applications*. John J. Grenfenstett (Ed.). Lawrence Erlbaum Associates: 285-306. Hillsdale, NJ.

GOLDBERG, D.E. (1989) *Genetic Algorithms in Search Optimization, and Machine Learning*. Addison Wesley. Reading, M.A.

GOLDBERG, D.E. Y P. SEGREST (1987) «Finite Markov chain analysis of genetic algorithms», *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms* John J. Grenfenstett (Ed.). Lawrence Erlbaum Associates: 1-8. Hillsdale, NJ.

GOLDBERG, D.E., B. KORB Y K. DEB (1989) «Messy Genetic Algorithms: Motivation, Analysis, and First Results», *Complex Systems*, 3: 493-530.

HERNÁNDEZ LÓPEZ, M. (2002) *Algoritmos Genéticos y Predicción de la Composición de la Demanda Turística*. Tesis Doctoral. Universidad de La Laguna.

HERNÁNDEZ LÓPEZ, M. (2004A) «Future tourists' characteristics and decisions: the use of genetic algorithms as a forecasting method», *Tourism Economics*, 10 (3): 245-262.

HERNÁNDEZ LÓPEZ, M. (2004B) «Predicción económica con algoritmos genéticos: operadores genéticos versus matriz de transición», *Estadística Española*, 46 (157).

HOLLAND, J.H. (1975) *Adaptation in Natural and Artificial Systems*. Ann Arbor. The University of Michigan Press.

HURLEY, S., L. MOUTINHO, Y S.F. WITT (1998), «Genetic algorithms for tourism marketing», *Annals of Tourism Research*, 25 (2): 498-514.

MAHFOUD, S. Y G. MANI (1996), «Financial forecasting using genetic algorithms», *Applied Artificial Intelligence*, 10: 543-565.

MICHALEWICZ, Z. (1994) *Genetic Algorithms + Data Structures = Evolution Programs*. Springer Verlag.

MORENO, J.M. Y J.A. MORENO (1999) *Heurísticas en Optimización*. Colección Textos Universitarios.

NIX, A. Y M.D. VOSE (1992) «Modeling genetic algorithm with Markov chains», *Annals of Mathematics and Artificial Intelligence*, 5: 79-88.

SUZUKI, J. (1995) «A Markov chain analysis on simple genetic algorithms», *IEEE Transactions on Systems, Man, and Cybernetics*, 25(4): 655-659.

VENKATESAN, R. Y V. KUMAR (2002) «A genetic algorithms approach to growth phase forecasting of wireless subscribers», *International Journal of Forecasting*, 18: 625-646.