UN ESTUDIO DE LA EVOLUCIÓN MIGRATORIA EN RAMAS DE ACTIVIDAD LABORAL MEDIANTE EL USO DE CADENAS DE MARKOV.

Josa Fombellida, Ricardo[1]
Dpto. de Economía Aplicada (Matemáticas)
Universidad de Valladolid

Abstract

En este trabajo se realiza un estudio de la evolución de las migraciones interiores entre distintas ramas de actividad laboral en los años 1993-97. Se pretende predecir, a corto y a largo plazo, la proporción de migrantes ocupados en cada categoría laboral. Para ello se modeliza el problema con una cadena de Markov finita y homogénea cuyos estados son las ramas de actividad. Se estiman las probabilidades de transición con el método de máxima verosimilitud. Tras comprobar la homogeneidad, se hallan las probabilidades de alcanzar los estados en los próximos años, sus probabilidades límite, las probabilidades de paso de un estado a otro en un número exacto de años y los tiempos medios de paso entre estados.

Palabras clave: Cadena de Markov, método de máxima verosimilitud, migración, rama de actividad.

1. Introducción.

El conocimiento de la distribución de trabajadores en ramas de actividad es importante desde el punto de vista económico. Es posible hacer un estudio de la evolución de esta distribución porcentual con una modelización matemática adecuada.

Los modelos matemáticos pueden ser categorizados como probabilísticos o como deterministas. En muchas situaciones los modelos aleatorios son más adecuados y esta mejor representación está dada considerando una colección de variables aleatorias. Esa colección indexada por un parámetro continuo o discreto es un proceso estocástico.

En estadística aplicada a datos empíricos, una distribución de probabilidad teórica es ajustada para extraer más información de los datos. Si el ajuste es bueno, las propiedades del

[1] Este trabajo ha sido subvencionado por el Proyecto V12/99 de la Consejería de Educación y Cultura de la Junta de Castilla y León.

conjunto de datos pueden ser aproximadas por las de las distribución. De forma similar, supongamos que se observa que un proceso de la vida real tiene características de un proceso estocástico. El conocimiento del comportamiento del proceso en cuestión es deseable para comprender la situación real.

El proceso estocástico más sencillo es aquel de parámetro discreto con posibles valores en un conjunto finito, que verifica una propiedad llamada propiedad de Markov. En un lenguaje informal ésta dice que conocido el presente, el futuro y el pasado son independientes. Dicho de otra forma, el estado del sistema en un momento cualquiera sólo depende del momento inmediatamente anterior. Esta dependencia de primer orden es la generalización inmediata de la independencia de variables aleatorias.

Este proceso estocástico simple, llamado cadena de Markov finita, será objeto de estudio en el trabajo y se aplicarán sus propiedades a datos reales de movimiento migratorio entre las distintas categorías de empleo.

El trabajo está dividido en dos secciones. En la primera, se expone de un modo breve terminología y resultados sobre cadenas de Markov finitas. En la segunda, se aplican esos resultados teóricos para estudiar aspectos de la evolución migratoria española.

2. Cadenas de Markov finitas.

2.1 Terminología y resultados.

En este apartado se exponen conceptos y resultados conocidos de cadenas de Markov finitas. Las demostraciones se pueden encontrar en [1], [2], [4] y [6].

Definición 1.

Una *cadena de Markov* es una sucesión de variables aleatorias $\{X_n\}_{n=0,1,2,...}$ definidas en un espacio probabilístico (W, s, p), que toman valores en un conjunto finito o numerable S Ì Â, $X_n : W \to \Re$, $\forall n$, y que verifican la *propiedad de Markov*, es decir, $p(X_n = i_n / X_{n-1} = i_{n-1},..., X_0 = i_0) = p(X_n = i_n / X_{n-1} = i_{n-1})$, $\forall n=1,2,...$, $\forall i_0, i_1,..., i_n \in S$, con $p(X_{n-1} = i_{n-1},..., X_0 = i_0) > 0$.

Los elementos de S son los *estados* de la cadena de Markov. Si $X_n=i$, se dice que *la cadena está en el estado i*.

Se dice que la cadena de Markov es *finita* si S es finito.

Se dice que es *homogénea* si $\forall n,\ \forall i, j \in S$, la probabilidad $p(X_n = j / X_{n-1} = i)$ no depende de n. En este caso, se denota por p_{ij} a esta probabilidad, que es la *probabilidad de "pasar de i a j en un paso"*.

En lo que sigue suponemos que tenemos una cadena de Markov finita y homogénea. Así podemos suponer que $S = \{1, 2, ..., m\}$.

Las *probabilidades de transición* de la cadena de Markov son $\{p_{ij}\}_{i \in S}$ y la *matriz de transición* es $P = (p_{ij}) \in M_{m \times m}([0,1])$. Es una matriz estocástica, es decir, $p_{ij} \geq 0$, $\sum_{j=1}^{m} p_{ij} = 1, \quad \forall i, j \in S$.

Definición 2.

El vector de *probabilidades en el momento* n es $a^n = (a_1^n, a_2^n, ..., a_m^n)$, con $a_i^n = p(X_n = i)$, $\forall\ i \in S$, $\forall\ n$. El vector de *probabilidades iniciales* es a^0.

Se verifica: $a^n = a^0 P^n$, $\forall\ n$ (1).

Definición 3.

Dados $i, j \in S$ y $n \in \{1,2,...\}$ se define la *probabilidad de "pasar de pasar de i a j en n pasos o menos"* como $p_{ij}^{(n)} = p(X_n = j / X_0 = i)$.

Se puede demostrar que la probabilidad anterior es el elemento (i, j) de la matriz P^n: $p_{ij}^{(n)} = (P^n)_{ij}$.

Teorema 1.

Sea $S \subset \mathbb{R}$ finito con m elementos, $\{a_i^0\}_{i \in S}$ con $a_i^0 \geq 0, \quad \forall i \in S$ y $\sum_{i \in S} a_i^0 = 1$, y P una matriz estocástica. Entonces existe una cadena de Markov $\{X_n\}_{n=0,1,2,...}$ definida en un espacio probabilístico (Ω, σ, P) tal que $\{a_i^0\}_{i \in S}$ son sus probabilidades iniciales y P es su matriz de transición.

Este resultado garantiza no sólo la existencia de una cadena de Markov con un número elegido de estados, sino que con un vector de probabilidades iniciales y una matriz estocástica determinamos la cadena de Markov.

Definición 4.

Dados $i, j \in S$, la probabilidad de que partiendo del estado *i* se alcance el estado *j* por primera vez en el paso *n*, o bien, la *probabilidad de pasar de i a j exactamente en n pasos* se define como $f_{ij}^{(n)} = p(X_1 \neq j, ..., X_{n-1} \neq j, X_n = j / X_0 = i)$, $\forall n$, $f_{ij}^{(1)} = p_{ij}$.

Definición 5.

Dados $i, j \in S$, la probabilidad de que partiendo del estado *i* se alcance el estado *j*, o bien, la *probabilidad de pasar de i a j* es $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$.

Definición 6.

Se dice que $i \in S$ es un estado *persistente* si $f_{ii} = 1$, es decir, si se pasa por *i*, con toda seguridad se retorna a *i*.

Definición 7.

Sean $i, j \in S$ persistentes y $T_j = \inf\{k \geq 1 / X_k = j\}$. El *tiempo medio en llegar a j desde i*, o bien el número esperado de transiciones antes de llegar a *j* desde *i* es $m_{ij} = E(T_j / X_0 = i)$. A m_{jj} se le denomina *tiempo medio de retorno* al estado *j*.

Se puede comprobar que $m_{ij} = \sum_{n=1}^{\infty} n \; f_{ij}^{(n)}$, $\forall \; i, j \in S$.

Definición 8.

Sean $i, j \in S$ persistentes. *El número de visitas al estado j en n pasos, partiendo del estado i* es el número de veces de entre las *n* que la cadena ha pasado por el estado *j*, dado que inicialmente estaba en *i*. Lo denotaremos por $N_{ij}^{(n)}$.

Se comprueba que $EN_{ij}^{(n)} = \sum_{k=1}^{n} p_{ij}^{(k)}$.

Fijado $i \in S$ y teniendo en cuenta que P^n es estocástica $\forall n$, se tiene que $\sum_{j \in S} EN_{ij}^{(n)} = n$.

Definición 9.

Sea $i \in S$ persistente, el *tiempo de ocupación del estado i* es el número de pasos que la cadena permanece en i antes de ir otro estado. Lo denotaremos por a_i.

Se demuestra que dado un estado $i \in S$ persistente, el tiempo de ocupación del estado *i* es una variable aleatoria geométrica con ley $p(a_i = n) = p_{ii}^n (1 - p_{ii})$, n=1,2,... y esperanza

$$E(a_i) = \frac{p_{ii}}{1 - p_{ii}} \qquad (2).$$

Definición 10.

Se dice que una cadena de Markov es *irreducible* si $\forall\, i, j \in S$, $\exists r$ con $p_{ij}^{(r)} > 0$. Es decir, si es posible ir de un estado a otro en cierto número de pasos.

Dada una cadena de Markov finita, hay un grafo dirigido asociado con nodos los estados de S y conjunto de arcos $\{(i, j) / i, j \in S, p_{ij} > 0\}$. Teniendo en cuenta que $p_{ij}^{(r)} = \sum_{i_1,\dots,i_{r-1} \in S} p_{ii_1} p_{i_1 i_2} \dots p_{i_{r-1} j}$, $\forall i, j \in S$, se puede pasar de *i* a *j* si es posible en el grafo "siguiendo un camino dirigido". Se puede ver si la cadena es irreducible, haciendo esta comprobación para cada par de estados.

Por otra parte, se puede demostrar que si una cadena de Markov es irreducible, todos sus estados son persistentes.

Definición 11.

Se dice que el estado $j \in S$ es *aperiódico* si m.c.d.$\{n \geq 1 / p_{jj}^{(n)} > 0\} = 1$.

Es inmediato que si $p_{jj} > 0$, para $j \in S$, entonces *j* es aperiódico.

Definición 12.

Una cadena de Markov es *aperiódica* si $\forall\, j \in S$, *j* es aperiódico.

Se puede comprobar que si una cadena de Markov es irreducible y tiene un estado $j \in S$ aperiódico, entonces la cadena de Markov es aperiódica (3).

Definición 13.

Se dice que $\overline{u} = (u_1\ u_2\ ...\ u_m)$ es una *distribución estacionaria* para *P* o para la cadena de Markov si: $u_i \geq 0, \forall i \in S, \quad \sum_{i=1}^{m} u_i = 1 \quad$ y $\quad \overline{u} = \overline{u}P \qquad$ (4).

Los siguientes resultados nos permiten el cálculo de los tiempos medios de paso entre estados por medio de la distribución estacionaria. También damos un resultado de estabilidad en la distribución estacionaria.

Proposición 1.

Si la cadena de Markov es irreducible y aperiódica, entonces existe una única distribución estacionaria $\overline{u} = (u_j)_{j \in S}$ y se verifica:

$$0 < u_j < 1, \qquad m_{jj} = \frac{1}{u_j} \qquad \text{y} \qquad \lim_{n \to \infty} p_{ij}^{(n)} = u_j, \quad \forall i, j \in S.$$

Proposición 2.

Si la cadena de Markov es irreducible los tiempos medios son solución de los sistemas de ecuaciones: $m_{ij} = 1 + \sum_{k \neq j}^{m} p_{ik} m_{kj}\ , \quad i, j \in S\ ,\ i \neq j\ .$

Proposición 3.

Dados $i, j \in S$ persistentes, entonces $\lim_{n \to \infty} \frac{EN_{ij}^{(n)}}{n} = u_j$. Es decir, el porcentaje de tiempo que se visita *j* en *n* pasos desde *i* se estabiliza en la *j*-ésima componente estacionaria.

En consecuencia, una cadena de Markov queda determinada por a^0 y *P*. Las probabilidades en el momento *n* se pueden hallar como se indica en (1):

$$p(X_n = i) = (a^0 P^n)_i = \sum_{k=1}^{m} a_k^0 (P^n)_{ki}\ .$$

En ese caso, si algún $p_{jj} > 0$, se tiene que la cadena es aperiódica, por (3). Una vez comprobado que la cadena de Markov es irreducible y aperiódica, si se utiliza la Proposición 1, se tienen las probabilidades límite: $\lim\limits_{n\to\infty} p(X_n = i) = \sum_{k=1}^{m} a_k^0 \lim\limits_{n\to\infty} p_{ij}^{(n)} = \sum_{k=1}^{m} a_k^0 u_i = u_i$ (5).

Para concluir el estudio, se pueden encontrar los tiempos medios de paso entre estados (Proposiciones 1 y 2) y los tiempos medios de permanencia en un estado antes de visitar otro (2).

2.2. Inferencia.

En esta sección consideraremos la estimación y contraste de hipótesis relativos a cadenas de Markov finitas y homogéneas. Sólo daremos algunos resultados de estimación máximo verosímil de los elementos de la matriz de transición y tests de hipótesis basados en ellos.

Hay varias cuestiones estadísticas importantes en el contexto de cadenas de Markov. En primer lugar, interesa determinar si la cadena de Markov es homogénea. En segundo lugar, asumiendo que el sistema puede ser razonablemente modelizado como una cadena de Markov homogénea, hay que estimar las probabilidades de transición. Finalmente, se pueden estudiar las características del sistema, como la distribución límite y su evolución en el tiempo.

Supongamos que observamos una cadena de Markov $\{X_n\}_{n=0,1,2,...}$ con conjunto de estados $S=\{1,2,\ldots,m\}$ hasta n transiciones. Es decir, en un intervalo de tiempo [0, T]. Sea n_{ij} el número de transiciones del estado i al estado j (i, $j \in S$). Sean $n_{i.} = \sum_{j=1}^{m} n_{ij}$ y $n_{.j} = \sum_{i=1}^{m} n_{ij}$, $i,\ j \in S$. Así, $n = \sum_{i=1}^{m} n_{i.} = \sum_{j=1}^{m} n_{.j}$.

Sea $P = (p_{ij})$ la matriz de transición. Estamos interesados en hallar los estimadores de sus elementos p_{ij}, que denotaremos por $\hat{p}_{ij}$ ($i, j \in S$).

Para cada $i \in S$ y $n_{i.}$, la muestra $(n_{i1}, n_{i2}, ..., n_{im})$ tiene una distribución multinomial con probabilidades $(p_{i1}, p_{i2}, ..., p_{im})$ tales que $\sum_{j=1}^{m} p_{ij} = 1$. Así, la función de verosimilitud es de la forma $f(P) = A \prod_{i=1}^{m} \frac{n_{i.}!}{n_{i1}! n_{i2}! ... n_{im}!} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} ... p_{im}^{n_{im}}$, y su logaritmo $L(P) = \ln B + \sum_{i=1}^{m} \sum_{j=1}^{m} n_{ij} \ln p_{ij}$, siendo B independiente de P.

La solución del problema $\max\{L(P)/\sum_{j=1}^{m} p_{ij} = 1, \quad \forall i \in S\}$ es el estimador máximo verosímil (EMV) de *P*. Se deriva respecto a p_{ij} ($\forall\, i,\, j \in S$) y se obtienen los estimadores $\hat{p}_{ij} = \dfrac{n_{ij}}{n_{i.}}$, $i,\, j \in S$.

Se puede probar que asintóticamente (cuando $n \to \infty$) la distribución de $\sqrt{n_{i.}}\,(\hat{p}_{ij} - p_{ij})$ es $N(0, p_{ij}(1 - p_{ij}))$. Basándonos en esta propiedad, se obtienen los estadísticos que contrastan hipótesis de interés. Este estudio lo hacemos a continuación.

En primer lugar, nos ocupamos de contrastar la hipótesis de que la cadena de Markov es homogénea.

Para cada *t*=1,2,...,*T* y $\forall\, i,\, j \in S$ consideramos la probabilidad de transición $p_{ij}^{t} = p(X_{t+1} = j / X_t = i)$. Sea n_{ij}^{t} el número de transiciones de *i* a *j* durante el período [*t*-1, *t*]. Argumentando como antes, el EMV de p_{ij}^{t} se puede obtener como $\hat{p}_{ij}^{t} = \dfrac{n_{ij}^{t}}{n_{i.}^{t}}$, donde $n_{i.}^{t} = \sum_{j=1}^{m} n_{ij}^{t}$.

Se desea contrastar la hipótesis $H_0 : p_{ij}^{t} = p_{ij}$, $\forall t$=1,2,...,*T*. El estadístico de razón de verosimilitudes es $L = \dfrac{f(\hat{P}^{t})}{f(P)}$. Bajo H_0, $-2\ln L$ tiene una distribución χ^2 con (*T*-1)*m*(*m*-1) grados de libertad. Se tiene que $-2\ln L = 2(L(\hat{P}^{t}) - L(P)) = 2\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m} n_{ij}^{t} \ln \dfrac{n_{ij}^{t}}{n_{i.}^{t}\, p_{ij}}$ (6).

Para contrastar la homogeneidad de $\{X_n\}_{n=0,1,2,...}$ en los periodos [0, 1], [1,2], ..., [*T*-1,*T*] ; hay que tomar los EMV de *P* en todo el periodo [0,*T*], es decir,

$$p_{ij} = \frac{\sum_{t=1}^{T} n_{ij}^{t}}{\sum_{t=1}^{T} n_{i.}^{t}}, \quad \forall\, i,\, j \in S \qquad (7).$$

A continuación contrastamos la hipótesis nula de que las observaciones son independientes contra la alternativa de que el proceso es una cadena de Markov. La hipótesis H_0 es que *P* tiene *m* filas iguales de la forma $p = (p_1, p_2, ..., p_m)$. La log-verosimilitud de *P* es $L(p) = \ln B + \sum_{j=1}^{m} n_{.j} \ln p_j$. Se obtienen los EMV $\hat{p}_j = \frac{n_{.j}}{n}$, $\forall j \in S$. Siguiendo razonamientos anteriores, el test de razón de verosimilitudes que contrasta esta hipótesis es

$$2\sum_{i=1}^{m}\sum_{j=1}^{m} n_{ij} \ln\left(\frac{n_{ij}n}{n_{i.}n_{.j}}\right) \qquad (8).$$

Bajo H_0, este estadístico sigue una distribución $c^2_{(m-1)^2}$.

En todos los tests anteriores, si para algún término $n_{ij} = 0$, la suma se extiende sólo a los sumandos con $n_{ij} \neq 0$ y los grados de libertad son los mismos menos el número de "ceros".

En resumen, conociendo el número de transiciones entre cada estado y para cada periodo, que están recogidos en las matrices $N^t = (n_{ij}^t)$, t = 1,2,...,T, es posible estimar la matriz de transición y contrastar que es una cadena de Markov homogénea.

3. Aplicación al estudio de la evolución de las migraciones interiores.

Utilizaremos los resultados teóricos anteriores para estudiar la evolución migratoria española. Previamente explicaremos brevemente la terminología sobre migraciones, para posteriormente concluir con la aplicación de lo comentado anteriormente a datos reales.

3.1. Conceptos referentes a la migración.

Toda la terminología que aparece a continuación esta recogida de [5].

Se considera *migrante* a toda persona que residiendo actualmente en un municipio, residía en la misma fecha del año anterior en otro municipio o bien en otro país. La *migración* es *interior* cuando los lugares de procedencia y destino corresponden a un mismo país.

El traslado de un grupo humano de una vivienda a otra dentro de la misma localidad carece de interés al estudiar los fenómenos generales de la migración. Por este motivo, sólo se consideran migraciones interiores a las que tienen lugar entre distintos términos municipales. Por tanto, se puede decir que *migrante interior* es toda persona que un periodo de referencia dado ha cambiado de municipio de residencia.

Se manejan los conceptos de *intra* e *inter* para distinguir entre migraciones que se efectúan dentro de una determinada área (intro) y las que se realizan dentro de áreas diferentes (inter). Se estudian los movimientos migratorios intra e inter referidos a provincias, comunidades autónomas, zonas que agrupan comunidades y tipos de hábitat. Estas zonas que agrupan comunidades son: Noroeste (Galicia, Asturias y Cantabria), Nordeste (País Vasco, Navarra, La Rioja y Aragón), Madrid, Centro (Castilla y León, Castilla - La Mancha y Extremadura), Este (Cataluña, Comunidad Valenciana y Baleares), Sur (Andalucía, Murcia, Ceuta y Melilla), Canarias y Extranjero (migraciones procedentes del extranjero). Hay dos tipos de *hábitat*: rural (municipio de menos de 20000 habitantes) y urbano (más).

Una ventaja de la Encuesta de Migraciones en relación a otras fuentes estadísticas que analizan el fenómeno migratorio es que proporciona información sobre las características de actividad de los migrantes. Es interesante estudiar las tasas de actividad y paro, así como su comparación con las de la población general. También es importante analizar la movilidad de los migrantes respecto a las *situaciones de actividad* (ocupado, parado, inactivo y población contada aparte (P.C.A.)) y compararlo con la movilidad de la población. Los significados de estas situaciones son: ocupado = trabajando, parado = buscando empleo, inactivo = disponible y sin buscar empleo, estudiando y otra situación; y P.A.C. = haciendo el servicio militar.

Si sólo analizamos a los *migrantes ocupados*, nos interesan las variables: situación profesional, ocupación y rama de actividad.

Nuestro trabajo se va a centrar en el colectivo de migrantes ocupados en función de las *ramas de actividad*:

ramas A/B: agricultura, ganadería, caza, selvicultura y pesca;

ramas C/D/E: industrias extractivas, industrias manufactureras, producción y distribución de energía eléctrica, gas y agua;

rama F: construcción;

ramas G/H: comercio y hostelería;

rama I: transporte, almacenamiento y comunicaciones;

ramas J/K: intermediación financiera, actividades inmobiliarias y de alquiler, servicios empresariales;

ramas L/M/N: administración pública, defensa y seguridad social, educación, actividades sanitarias y veterinarias;

ramas O/P/Q: otros servicios.

3.2. Análisis de resultados.

Se considera a los migrantes ocupados en la actualidad en el municipio de destino que también trabajaban el año anterior en su lugar de origen. En [5], se dispone de su distribución en función de las ramas de actividad. Es decir, conocemos el movimiento de migrantes ocupados de una a otra rama de actividad, de un año a otro. Esta tabla de contingencia se tiene para los años 1992/93, 1993/94, 1994/95, 1995/96 y 1996/97. Asimismo también se conoce la proporción de migrantes ocupados dentro de cada rama para los años 1993-97.

Modelizamos esto con una cadena de Markov finita, con conjunto de estados S={ramas A/B, ramas C/D/E, rama F, ramas G/H, rama I, ramas J/K, ramas L/M/N, ramas O/P/Q} y matriz de transición P que estimaremos.

Con muestras notaciones conocemos las matrices N^1, N^2, N^3, N^4 y N^5. El vector de proporciones de cualquier año se puede tomar como vector de probabilidades iniciales a^0.

Utilizamos el programa MATLAB para hacer los cálculos, obteniéndose lo siguiente.

El estadístico de razón de verosimilitudes (6) toma el valor $-2\ln\Lambda=39.33$. Bajo la hipótesis nula de homogeneidad tiene una distribución χ^2 con $3\times8\times7-197=27$ grados de libertad. En las tablas χ^2 encontramos que $P(\chi^2>39.33)>0.04$, con lo cual no se puede rechazar la hipótesis de que la cadena de Markov es homogénea, a un nivel de confianza del 1%.

El estadístico de razón de verosimilitudes (8) toma el valor $-2\ln\Lambda=546.3383$. Bajo la hipótesis nula de independencia tiene una distribución χ^2 con $7^2-15=34$ grados de libertad. Se tiene que $P(\chi^2>546.3383)<0.0001$, y por tanto, se puede rechazar la hipótesis de que las observaciones son independientes.

Se acepta que los datos se pueden ajustar por una cadena de Markov finita y homogénea de primer orden.

La matriz de transición se calcula como en (7) y está dada abajo:

$$P=\begin{pmatrix} 0.6842 & 0.0947 & 0.1053 & 0.0842 & 0 & 0.0316 & 0 & 0 \\ 0.0314 & 0.8431 & 0.0039 & 0.0431 & 0 & 0.0314 & 0.0275 & 0.0196 \\ 0.0650 & 0.0200 & 0.8600 & 0.0350 & 0 & 0.0100 & 0.0100 & 0 \\ 0.0309 & 0.0365 & 0.0140 & 0.8483 & 0.0112 & 0.0253 & 0.0253 & 0.0169 \\ 0 & 0.0297 & 0.0693 & 0 & 0.8515 & 0 & 0.0396 & 0.0099 \\ 0 & 0.0222 & 0.0167 & 0.0278 & 0 & 0.9222 & 0.0111 & 0 \\ 0.0052 & 0.0017 & 0.0034 & 0.0172 & 0 & 0 & 0.9691 & 0.0034 \\ 0.0116 & 0.0058 & 0 & 0.581 & 0.0116 & 0.0116 & 0.0174 & 0.8837 \end{pmatrix}$$

Todos sus elementos diagonales son positivos, luego la cadena de Markov es aperiódica. También es fácil ver que es irreducible.

Se observa que de un año a otro es mayor la probabilidad de mantenerse en la misma rama de actividad, que la de pasar a otra rama.

Tomamos como vector de probabilidades iniciales el de las proporciones medias en cada rama de los años 93-97. Se obtiene a^0 = (0.0543, 0.1301, 0.1064, **0.2009**, 0.0448, 0.0890, **0.2733**, 0.0989).

Es decir, los migrantes ocupados trabajan más en las ramas G/H (comercio y hostelería) y L/M/N (administración pública) con porcentajes del 20.09 y del 27.33 por ciento, respectivamente.

Para el año 1998 se esperan proporciones similares: $a^1 = a^0 P$ = (0.0569, 0.1287, 0.1061, 0.1973, 0.0415, 0.0951, 0.2774, 0.0947).

A medio plazo, pongamos por ejemplo dentro de 5 años, se observa que se mantiene la importancia de las ramas pública (29.05%) y hostelera (18.82%):

a^5 = (0.0602, 0.1254, 0.1055, **0.1882**, 0.0320, 0.1144, **0.2905**, 0.0815).

Podemos predecir la proporción de migrantes ocupados en cada rama en el futuro (a largo plazo). Resolviendo (4) y usando (5) resulta: $\lim_{n\to\infty} a^n = \bar{u}$ =(0.0561, 0.1165, 0.0972, **0.1733**, 0.0175, **0.1470**, **0.3362**, 0.0562).

Observamos que en relación con las proporciones medias 93-97, se producirá un aumento en las ramas J/K (intermediación financiera) y L/M/N (administración pública), que junto con la G/H (comercio y hostelería) serán las más importantes.

Utilizando las proposiciones 1 y 2 se tienen los tiempos medios de paso entre las distintas ramas:

	A/B	**C/D/E**	**F**	**G/H**	**I**	**J/K**	**L/M/N**	**O/P/Q**
A/B	**17.8253**	33.2951	53.5606	25.1047	373.4509	66.3493	66.9619	147.2277
C/D/E	58.8513	**8.5825**	78.7084	29.9767	374.3690	68.2275	59.7891	136.5414
F	40.3520	41.8655	**10.2835**	28.6665	377.3140	72.4324	67.2562	151.9752
G/H	56.7515	42.8301	72.4342	**5.7719**	352.3859	69.6537	62.8430	138.0187
I	62.6660	49.1440	53.1941	39.6951	**57.1211**	86.6855	54.1929	147.1504
J/K	69.8253	47.4638	74.7407	33.8830	382.3104	**6.8031**	66.8063	156.3256
L/M/N	78.7695	72.3296	94.9399	44.3138	390.5664	102.5401	**2.9745**	158.1988
O/P/Q	64.8847	53.7733	81.0166	26.5662	337.1251	77.4033	61.7223	**17.798**

En la diagonal aparecen los tiempos medios de retorno.

Así, si un trabajador ocupado en la rama L/M/N (administración pública) cambia a un empleo de otra rama, tardaría 2.97 años en volver a trabajar en ésta. Uno que trabaja en I (transporte) deja ésta para no volver (57.12 años).

Los tiempos medios de ocupación de cada estado son Eα=(2.1666, 5.3735, 6.1429, 5.5920, **5.7340**, 11.8535, **31.3625**, 7.5985).
Se espera que un migrante trabajando en la rama I esté en ésta 5.73 años antes de cambiar a otra rama. Un trabajador de las ramas L/M/N se jubilaría en ella (31.36 años).

Las probabilidades de paso en un número exacto de años (definición 4) son relativamente pequeñas y en este ejemplo no tienen mucho interés.

Las probabilidades de cambiar de una rama de actividad a otra en 5 años o menos se calculan como $\sum_{i=1}^{5} p(a_i = n)$ y están dadas a continuación:

A/B	C/D/E	F	G/H	I	J/K	L/M/N	O/P/Q
0.5816	0.4839	0.4554	0.4756	0.4704	0.3070	0.1409	**0.4074**

Así, un migrante que trabaja en la rama O/P/Q (otros servicios) cambia a una ocupación de otra rama, en 5 años como mucho, con una probabilidad de 0.4.

De cada 10 años, por término medio, un trabajador que está en la rama *i* actualmente trabaja el siguiente número de años en cada rama *j*. Estos elementos (*i*, *j*) de $EN^{(10)} = \sum_{k=1}^{10} P^k$ están dados en la tabla:

	A/B	**C/D/E**	**F**	**G/H**	**I**	**J/K**	**L/M/N**	**O/P/Q**
A/B	**2.6303**	1.8401	1.8789	1.8728	0.0647	1.0614	0.4378	0.2140
C/D/E	0.6212	4.7936	0.4032	1.3781	0.0586	1.0964	1.0768	0.5721
F	1.1754	0.9608	5.2532	1.3125	0.0400	0.6156	0.5280	0.1144
G/H	0.6430	1.1310	0.6522	**5.0295**	0.2808	0.9497	0.7993	0.5145
I	0.3266	0.8722	1.7352	0.4489	**4.6042**	0.2004	1.4869	0.3255
J/K	0.1821	0.7787	0.5788	0.9866	0.0301	**6.7574**	0.5944	0.0919
L/M/N	0.1874	0.1737	0.1964	0.6585	0.0229	0.0877	**8.5126**	0.1608
O/P/Q	0.3538	0.4608	0.2621	1.6662	0.3443	0.5690	0.8326	**5.5112**

Se ve que cada migrante ocupado pasa la mayoría de los 10 años en la misma rama, en la que estaba inicialmente. En particular, uno de las ramas L/M/N está una media de 8.51 años de los 10 en esta misma rama.

4. Conclusiones.

Aprovechando que se conocen el número de transiciones entre distintas categorías (ramas de actividad) de un año a otro, hemos modelizado el sistema como una cadena de Markov finita con estados las categorías. Resultó ser homogénea, irreducible y aperiódica. Se conocen resultados de este tipo de procesos que permiten estudiar la evolución en el tiempo a través del cálculo de ciertos vectores de probabilidades. Utilizando todo eso se han calculado las proporciones a medio y largo plazo en cada categoría. Se observa que se mantiene la importancia actual de las ramas pública, comercial y hostelera en el futuro, tomando importancia la rama de intermediación financiera. Un trabajador permanece en cada rama un periodo de tiempo relativamente pequeño, salvo para la rama pública.

Resultaría interesante estudiar, modelizando con una cadena de Markov, la evolución de las proporciones de migrantes en las cuatro situaciones de actividad y compararla con la de la población general. Asimismo se podría comparar los resultados que hemos obtenido con los de la población general. También tendría interés el estudio de la evolución del movimiento migratorio entre provincias, entre comunidades y entre zonas que agrupan comunidades.

Referencias.

[1] BHAT, U.N. (1984). *Elements of Applied Stochastic Processes*. John Wiley & Sons, Inc. New York.

[2] BILLINGSLEY, P. (1995). *Probability and Measure*. John Wiley & Sons. New York.

[3] HILLIER, F. and G. LIEBERMAN. (1991). *Introducción a la Investigación de Operaciones*. Mc Graw-Hill. México.

[4] ISAACSON, D. and R. MADSEN. (1976). *Markov Chains: Theory and Applications*. John Wiley & Sons, Inc. New York.

[5] INE. (1993-1997). *Encuesta de Población Activa. Encuesta de Migraciones*.

[6] KARLIN, S. and H.M. TAYLOR. (1975). *A First Course in Stochatic Processes*. Academic press. New York.

[7] WINSTON, W. (1994). *Investigación de Operaciones*. Grupo Editorial Iberoamérica. México.