

ANÁLISIS DEMOGRÁFICO DE LAS BIOGRAFÍAS

Daniel Courgeau y Éva Lelièvre

Traducción del francés de Mariela Álvarez



EL COLEGIO DE MÉXICO

ÍNDICE

Agradecimientos	11
Introducción	13
PRIMERA PARTE	
GENERALIZACIÓN DEL ANÁLISIS LONGITUDINAL	
I. La observación de biografías	21
A) Diversos tipos de encuestas	22
B) La encuesta "Triple biografía"	25
C) El problema de las ponderaciones	28
D) La observación incompleta y errónea de las encuestas	30
E) El problema de los truncamientos	36
F) Conclusión	38
II. Formalización del análisis	41
A) Análisis de una cohorte homogénea que experimenta un solo evento	42
1) Distribución de la ocurrencia de los eventos, notaciones y formalización	42
2) La función de verosimilitud	44
3) Tiempo discreto, tiempo continuo	45
B) Análisis de una cohorte heterogénea y de la interacción entre fenómenos	48
1) Demografía diferencial	49
2) Modelo markoviano o semimarkoviano	50
a) Formalización del modelo	51
b) Relación del modelo con los procesos de conteo poissonianos	53
c) Generalización del modelo	55
C) Hacia un análisis más completo de los comportamientos humanos	58
D) Conclusión	61
III. Métodos de estimación a partir de observaciones truncadas	63
A) Presentación de los truncamientos	63
B) Truncamientos a la derecha	65
1) Estimador de Kaplan-Meier	66
Ejemplos miniatura	70
2) Estimador de Aalen	72

C) Truncamientos a la izquierda	74
1) Un método de corrección	75
2) Resultados obtenidos	77
D) Conclusión	80
IV. Estudio de un evento	81
A) Muestra simple. Un evento único	81
1) Estimación en tiempo discreto	81
2) Estimación actuarial	82
3) Estimación de los cocientes acumulados	84
B) Muestra simple: riesgos múltiples	85
Ejemplos	87
C) Muestras múltiples: pruebas comparativas	87
Ejemplos	90
D) Conclusión	94
V. Estudio recíproco de las interacciones entre dos eventos	95
A) Concepción del análisis	95
1) Dos fenómenos en competencia	95
2) La diferenciación dinámica	96
3) Los diferentes tipos de dependencia	96
4) Dependencia unilateral y causalidad	99
B) Formalización del análisis bivariado	100
1) Estimación actuarial	102
2) Las pruebas posibles	103
3) Las pruebas seleccionadas y la convergencia de las estimaciones	104
C) Análisis práctico	106
1) Simultaneidades e intervalos de tiempo	108
2) Representaciones gráficas	110
D) Conclusión	111
VI. Generalización para situaciones más complejas	113
A) Presentación y límites de la aplicación práctica	113
B) Interacciones entre tres eventos: dos casos de estudio	114
C) Interacciones entre dos eventos con ocurrencias múltiples	118
D) Conclusión	120

SEGUNDA PARTE

GENERALIZACIÓN DE LOS MODELOS DE REGRESIÓN

VII. Formalización estadística del análisis paramétrico	123
A) Algunos modelos paramétricos útiles en demografía	123
1) Distribución exponencial	124
2) Mezcla de distribuciones exponenciales	126
3) Distribución de Gompertz	134

4) Distribución de Weibull	139
5) Distribución Gamma	141
6) Distribución log-normal	143
7) Distribución log-logística	144
8) Distribución de Fisher-Snedecor (F) generalizada	146
9) Comparación de las diversas distribuciones	148
B) Modelos de regresión	149
1) Modelos de riesgos proporcionales	150
a) Distribución exponencial	152
b) Distribución de Weibull	152
c) Distribución de Gompertz	153
d) Distribución log-logística	153
e) Verificación de la validez del modelo	153
2) Modelos de tiempo de ocurrencias aceleradas	156
a) Nexos entre los modelos de tiempo de ocurrencias aceleradas y los modelos de riesgos proporcionales	156
b) Distribución log-logística	157
c) Verificación de la validez del modelo	158
3) Modelos más complejos	159
C) Conclusión	159
VIII. Métodos de estimación de modelos paramétricos	161
A) Cálculo de la verosimilitud cuando ciertos intervalos están abiertos	162
B) Estimación de los parámetros y pruebas de su valor	164
C) Ejemplos de estimación de los parámetros	165
1) Modelo exponencial	165
2) Utilización del método de Newton-Raphson	173
3) Modelo de Weibull	174
4) Modelo de Gompertz	179
5) Modelo log-logístico de ocurrencias aceleradas	184
D) Comparación de modelos paramétricos	190
E) Conclusión	194
IX. Métodos de análisis semiparamétrico	197
A) De las regresiones paramétricas a los modelos de riesgos proporcionales semiparamétricos	197
1) Definición	197
2) Construcción de la verosimilitud	198
B) Métodos de estimación	199
1) Estimación de los parámetros	199
2) Estimación del componente no paramétrico	202
C) Algoritmo de Newton-Raphson	204
D) Selección de un modelo para el análisis de interacciones	205
E) Algunos ejemplos de aplicación	206

F) Conclusión	210
Conclusión general	211
A) El análisis de las interacciones entre fenómenos	211
1) Formalización teórica	212
2) Modelos no paramétricos	214
B) El tratamiento de la heterogeneidad de las poblaciones	216
1) Modelos paramétricos	217
2) Modelos semiparamétricos	218
C) Nuevas líneas de investigación	220
Anexo 1	
I. Correspondencia de los términos en inglés, francés y español	223
II. Fórmulas básicas	225
III. Recordatorios y demostraciones	227
Producto integral	227
Prueba de una hipótesis nula, prueba general	228
Normalidad asintótica	229
Relación entre los procesos de conteo y el análisis de regresión de los intervalos	230
Planteo de Cox para la estimación de la verosimilitud parcial	231
Proceso de conteo y martingalas (Aalen)	232
IV. <i>Softwares</i> disponibles	233
Anexo 2	241
Anexo 3	279
Bibliografía por temas	291

INTRODUCCIÓN

Hasta el presente, el análisis longitudinal básicamente se ha desarrollado considerando cada fenómeno demográfico por separado. El principal objetivo ha sido aislar cada uno de estos fenómenos en estado puro. En particular se ha considerado necesario separar la influencia del evento estudiado de la de los fenómenos perturbadores: la mortalidad y los movimientos migratorios. Para ello se ha planteado un cierto número de hipótesis, sin que realmente haya sido posible verificarlas con las fuentes existentes (Henry, 1959, 1966).

Hay que considerar que este análisis se ha desarrollado a partir de datos agregados, como las estadísticas del estado civil o, en el mejor de los casos, los datos de los registros de población. Si bien estas fuentes ofrecen la posibilidad de tratar cada evento eliminando el efecto perturbador de los otros, casi no permiten analizar las interacciones entre fenómenos. Por lo tanto, en los manuales de demografía clásica encontramos que los fenómenos están aislados en estado puro, y se tratan por separado en cada uno de los diferentes capítulos: nupcialidad, fecundidad, mortalidad, desplazamientos y migraciones (Pressat, 1961; Henry, 1972).

No obstante, algunos demógrafos han destacado la utilidad de analizar las interacciones entre los fenómenos demográficos. Así, Pressat (1966) hacía énfasis en que "la búsqueda de las correlaciones entre fenómenos demográficos, si bien todavía constituye un dominio inexplorado, debería poder enriquecer considerablemente nuestro conocimiento". Pressat, sin embargo, no indicaba ningún camino a seguir para el análisis de esas correlaciones. Asimismo, cuando Henry (1972) hizo un examen de la nupcialidad, indicaba que "en cuanto a los emigrantes, podríamos vernos tentados a sustituir su nupcialidad en el extranjero por la que habrían tenido si se hubieran quedado, siendo que tal nupcialidad en el extranjero depende de condiciones que pueden ser muy diferentes". Este reconocimiento de la interacción entre los dos fenómenos y el cambio de comportamiento respecto de la nupcialidad una vez que los individuos han emigrado, es de mucha validez. Sin embargo, ante la ausencia de datos, Henry no llevó más lejos su análisis.

Henry (1959) también abordó otro problema importante: la heterogeneidad. Si bien "en el caso de una cohorte homogénea, la historia estadística de los individuos que la componen es la misma que la historia estadística

de la cohorte", cuando trabajamos con una cohorte heterogénea ese resultado ya no se sostiene. Así, por ejemplo, en el caso más simple que es el de una población que puede descomponerse en dos subpoblaciones —cada una con una probabilidad diferente pero constante a lo largo del tiempo en que se experimenta el evento estudiado—, la población en conjunto no tendrá un cociente constante, esto es, la media de los cocientes de las dos subpoblaciones. Si tal es el caso en el instante inicial, al paso del tiempo se produce una selección que elimina de la población que está expuesta al riesgo a la subpoblación que tiene el cociente más elevado, lo que ocasiona que a partir de cierto tiempo el cociente de la población observada tienda hacia el de la subpoblación que tiene el cociente más bajo. Por supuesto que esta heterogeneidad podría ser mucho más compleja, y todo indica que es necesario estudiarla.

"Para saber exactamente cuál es el efecto de la heterogeneidad de los grupos humanos habrá que orientar las investigaciones de la demografía diferencial hacia las características individuales, físicas y psicológicas, preocupándose por estudiar a la vez la dispersión y la correlación de los índices demográficos en el interior de los grupos, bastante sucintos, hasta aquí considerados" (Henry, 1959).

Mientras el demógrafo siga utilizando las estadísticas del estado civil o de los registros de población que se publican habitualmente, no tendrá manera de abordar estos dos problemas fundamentales: el análisis de las interacciones entre los fenómenos demográficos y el análisis de la heterogeneidad de los grupos humanos. Por lo tanto, hay que valerse de otras fuentes que permitan una observación continua de un grupo de individuos a lo largo de toda su existencia, o al menos de una parte de ella, y de la recolección del mayor número de características en cada encuesta.

Vemos entonces que la unidad de análisis ya no será el evento (deceso, matrimonio, nacimiento, migración, etc.) sino la biografía individual, considerada como un proceso complejo. En la actualidad ya no sólo se pretende aislar cada fenómeno en estado puro sino, por el contrario, se intenta ver cómo un evento en una existencia puede influir sobre la continuación de la vida individual y cómo ciertas características pueden empujar a un individuo a que se comporte de manera diferente a otro.

Este cambio de perspectiva nos conduce a reformular las bases del análisis demográfico en términos del examen de procesos estocásticos complejos. Tratemos de ver con más precisión cómo llegar a eso.

Esos procesos no se producen en un espacio-tiempo abstracto, sino que tienen su fuente en una estructura social particular. Un individuo nacido en una tribu Lobi a comienzos del siglo xx tendrá una biografía de un tipo completamente diferente a la de un individuo nacido en la Francia campesina de la misma época, o a la de uno nacido en la Francia urbana actual. Sin

embargo, en cada una de esas estructuras sociales podemos distinguir sistemas de relaciones que están más o menos desarrollados según el grupo o la sociedad considerados (Kimball y Pearsall, 1954): sistemas familiares, económicos, políticos, religiosos, educativos, asociativos e informales. Claro que nada impide que en el futuro aparezcan nuevos tipos de sistemas de relaciones. Nuestro acercamiento no considera a la sociedad como algo cerrado sino, al contrario, en evolución perpetua.

Cada miembro de una sociedad dada está simultáneamente implicado en los diversos sistemas. Así, un individuo que vive en la actualidad en Francia puede estar implicado en el sistema familiar como cohabitante y padre de un niño; en el sistema económico como ingeniero en la industria automotriz; en el sistema político como consejero municipal de su comuna; en el sistema religioso como católico no practicante; en el sistema educativo como alguien que continúa con cursos de perfeccionamiento de su especialidad; en el sistema asociativo como miembro de un equipo de fútbol no profesional y, por último, en el sistema informal como participante ocasional en las reuniones de padres de alumnos, con el objeto de resolver los problemas de la educación de su hijo.

Es la interacción entre esos diversos tipos de participaciones lo que va a generar un espacio y un tiempo propios para cada situación. La movilidad espacial o profesional de un soltero quizás sea más frecuente y cubra distancias más grandes que la de un individuo casado, sobre todo si éste tiene uno o varios hijos. En efecto, este último está ligado a su lugar de residencia y a su trabajo por restricciones debidas al lugar de trabajo de su esposa, al lugar donde van a la escuela sus hijos, etcétera.

El análisis de biografías pretende entonces situar esos cambios en el tiempo y el espacio vividos por los individuos en el marco de su sociedad. De lo que se trata es de ver cómo un acontecimiento familiar, económico o de otro tipo que enfrenta un individuo, modificará la probabilidad de que se produzcan otros eventos en su existencia. Se tratará de ver, por ejemplo, cómo su matrimonio puede influir en su carrera profesional, su movilidad espacial, la aparición de otros eventos como el nacimiento de un niño, o la ruptura con su familia de origen, etcétera.

Llegamos así al análisis de las interacciones entre fenómenos demográficos que tanto nos interesa. El área de estudio de este análisis es la de las biografías individuales.

Si tratamos de comprender los comportamientos de un individuo habremos de interesarnos por sus orígenes sociales y toda su historia pasada. Aquí partimos del supuesto de que esos comportamientos no son innatos sino que se modifican a lo largo de la existencia individual, gracias a las experiencias personales y a sucesivas adquisiciones. Así, dos individuos con el mismo origen social pero que hayan seguido caminos completamente

distintos, tendrán un comportamiento frente al matrimonio, la constitución de la familia, la carrera profesional, etc. que variará a lo largo del tiempo.

Con esto llegamos al análisis de la heterogeneidad de las poblaciones, aunque visto en forma dinámica y no estática. Este análisis, que ocupa un lugar distinguido en el estudio de las biografías individuales, no es determinista sino probabilista desde sus inicios.

Así pues, varios individuos colocados en una misma situación conflictiva tendrán desde el principio probabilidades diferentes de encontrar soluciones a esta situación antes de una cierta fecha. Algunos quizás no las encuentren jamás, otros podrán inventar un comportamiento totalmente nuevo que quizás resuelva mejor la situación de conflicto en la que se encontraban. Le dejamos así al individuo un margen de libertad que puede conducir a situaciones completamente nuevas. Este margen de libertad permite una evolución de la humanidad no inscrita en el origen, sino desarrollada de forma irreversible a lo largo del tiempo.

En este sentido nos encontramos muy cerca de Prigogine y Stengers (1988), quienes reconocían que "el evento crea una diferencia entre el pasado y el futuro... es el producto inteligible de un pasado del que, sin embargo, no podría deducirse. Abre hacia un futuro histórico donde se decidirá si sus consecuencias son insignificantes o tienen sentido". Encontramos aquí ese margen de libertad que puede conducir a situaciones completamente nuevas.

¿Podemos entonces formalizar este análisis de las biografías que inicialmente presentamos de manera informal?

Cuando un individuo nace, su existencia puede seguir una gran variedad de trayectorias. A pesar de ello, esas diversas trayectorias distan de ser todas igualmente probables. La biografía de un individuo puede definirse entonces como el resultado de un proceso estocástico complejo que se desarrolla a lo largo del tiempo, pero que se sitúa en condiciones históricas, geográficas, económicas y sociales dadas.

Llamemos Ω_θ al conjunto de las historias de vida o partes de historias de vida que podríamos observar hasta un instante θ . Como ya dijimos, debemos limitarnos a la observación del pasado, pues el futuro introduce situaciones nuevas no deducibles del pasado. Por ejemplo, antes de la aparición del automóvil no se podía prever que surgiera la profesión de mecánico, y quizás en un futuro cercano los descubrimientos genéticos extenderán la duración de la vida humana hasta los 150 años o más. El análisis que se puede realizar en el instante t no toma en cuenta más que los comportamientos pasados, y los proyecta hacia el porvenir sin tomar en cuenta su evolución. Ésta, sin embargo, se produce a un ritmo lo bastante lento como para que el análisis del pasado esclarezca las probabilidades de los diversos eventos del futuro próximo.

Sea ω_n una biografía enteramente observada, donde n precede a θ y ω_θ una biografía observada hasta el instante θ , que no está terminada. Se puede decir que los eventos de una u otra de esas biografías son variables definidas sobre el espacio general Ω_θ . Por ejemplo, la edad en la que se casa uno de esos individuos es una aplicación (medida de probabilidad) de Ω_θ sobre la abscisa positiva $([0, +\infty])$.

Entonces dotemos a Ω_θ de una sigma-álgebra¹ $\mathcal{B}_\theta$ de los eventos que deseamos analizar en una población dada y de una aplicación (medida de probabilidad) P_θ de $\mathcal{B}_\theta$ sobre $(0, 1)$, que nos va a indicar las probabilidades de conocer esos diferentes eventos en la población estudiada. En ese caso $(\Omega_\theta, \mathcal{B}_\theta, P_\theta)$ define bien un campo de probabilidad.

Un instante aleatorio T será entonces una función de tiempo en el espacio de probabilidad de Ω_θ , que en este caso se puede extender más allá de θ , suponiendo que las probabilidades definidas antes de θ permanezcan las mismas en el curso del tiempo. Así, por ejemplo, si un individuo no está casado en θ , se supone que su edad al casarse es una variable aleatoria T que sigue la misma distribución que la observada para los individuos ya casados con características semejantes a las de él. Esta hipótesis ni siquiera es necesaria, y podemos contentarnos con trabajar sobre las biografías completamente terminadas, como se hace en la demografía histórica.

El análisis de las biografías que proponemos aquí consistirá, pues, en estimar la distribución de las probabilidades de las trayectorias seguidas por una población dada. Esta distribución puede variar de una subpoblación a otra y depender de ciertas cualidades de los individuos de la subpoblación (características sociales y económicas de los padres y de los abuelos, por ejemplo). Estas trayectorias se identifican por variables aleatorias $T_1, T_2, \dots, T_n$, que representan las duraciones de permanencia en los diversos estados que las componen. Claro está que esas variables no son independientes y la distribución de las trayectorias que queremos estimar resulta de su distribución conjunta.

Nuestro acercamiento supone, pues, que el comportamiento de los individuos se puede describir como un proceso estocástico complejo. Una vez admitido ese modelo, se comienza por la estimación estadística de la distribución de las variables anteriormente definidas, que desarrollamos en esta obra. Habiendo ya conocido esas distribuciones, será posible acercarse a la distribución más compleja del conjunto de la trayectoria.

Partiremos ante todo de los métodos para conseguir esas biografías. Por lo general no se puede disponer de una recolección exhaustiva de las biografías, sino de datos recogidos de una muestra de individuos cuya biografía

¹ Conjunto de las partes Ω_θ .

completa no se observa. Por lo tanto, tendremos que evaluar los diversos problemas que nos plantea esta observación incompleta.

A continuación formalizaremos el análisis y los métodos de estimación partiendo de casos simples para irlos complicando progresivamente. Del estudio de un evento pasaremos al estudio recíproco de las interacciones entre dos eventos, antes de generalizar esos modelos no paramétricos hacia situaciones más complejas. Pasaremos luego a los modelos paramétricos, los cuales permiten hacer que intervenga el efecto de un gran número de características sobre la duración de la permanencia en un estado dado. Por último, introduciremos modelos semiparamétricos que combinan los dos acercamientos precedentes.

Estos diversos métodos se ilustrarán mediante su aplicación a situaciones muy diferentes con la intención de mostrar su generalidad. Como anexo se incluyen ciertos programas que permiten realizar esos análisis, a fin de que el lector del libro pueda utilizarlos.

Este manual suministra, a la vez, una presentación teórica detallada de los métodos de análisis de las biografías y una posibilidad de aplicación práctica de esos métodos tanto a los archivos ya existentes, como a aquellos que se creen a partir de encuestas biográficas.

CENTRO DE ESTUDIOS DEMOGRÁFICOS
Y DE DESARROLLO URBANO

Durante los últimos años, las técnicas de análisis demográfico han alcanzado un desarrollo importante, tal es el caso del análisis de historia de eventos. Incluso esta metodología ha sido utilizada ampliamente desde los ochenta en áreas tales como la economía y la bioestadística.

Para la demografía, el análisis de la historia de eventos amplía el alcance del enfoque longitudinal tradicional y provee resultados adicionales que son esenciales para el entendimiento del comportamiento humano en toda su complejidad.

El análisis de historia de eventos tiene su origen en el análisis longitudinal tradicional. Después de la segunda guerra mundial, el análisis de cohortes reemplazó al análisis transversal, el cual se utilizaba desde fines del siglo XVII. El análisis longitudinal estudia la ocurrencia de un evento demográfico (matrimonio, nacimiento, defunción, migración, etc.) dentro de un determinado grupo homogéneo o cohorte.

Por otra parte, el análisis de historia de eventos estudia la aparición de uno o más fenómenos demográficos interrelacionados. Puede involucrar simultáneamente a un gran número de características individuales, algunas de las cuales podrían cambiar con el tiempo. Esto proporciona una adecuada solución a algunos de los problemas que se presentan en el análisis longitudinal tradicional.

ISBN 9-681-20968-0



Ambassade de France - CCC IFAL

Centre Culturel et de Coopération