

1

Introducción

1.1 Motivación

Cuando escribí mi primer libro, *“Qualitative Choice Analysis”*, a mediados de los años 80, este campo del conocimiento había alcanzado un momento crítico. Los conceptos innovadores que lo definían habían sido descubiertos. Los modelos básicos – principalmente logit y logit jerárquico – habían sido introducidos, y las propiedades estadísticas y económicas de estos modelos se habían inferido. Estos conceptos habían sido aplicados con éxito en diferentes áreas, incluyendo transporte, energía, vivienda y marketing, por nombrar sólo unas cuantas.

Este ámbito está hoy en día en un momento similar en relación a una nueva generación de procedimientos. Los modelos de primera generación tenían limitaciones importantes que reducían su utilidad práctica y su realismo. Esas limitaciones fueron claramente identificadas en su momento, pero la forma de superarlas no había sido descubierta. A lo largo de los últimos veinte años se han realizado enormes progresos, lo que nos ha llevado a un cambio radical en los métodos de análisis de la elección. Los primeros modelos han sido complementados por nuevos métodos, más potentes y flexibles. Los nuevos conceptos han surgido gradualmente, gracias a investigadores edificando sobre el trabajo de otros investigadores. Sin embargo, en cierto modo, el cambio ha sido más parecido a un salto brusco que a una progresión gradual. La forma en que los investigadores piensan, especifican y estiman sus modelos, ha cambiado. Y lo que es más importante, un alto grado de consenso, o de comprensión, parece haber emergido en relación a la nueva metodología. Entre los investigadores que trabajan en este campo, un evidente sentido del propósito y del progreso prevalece.

Mi propósito al escribir este nuevo libro es reunir todas estas ideas, en una forma que ejemplifique la unificación de criterios que a mi parecer se ha logrado, y de una manera que haga estos métodos accesibles para una amplia audiencia. Los avances se han centrado principalmente en la simulación. En esencia, la simulación es la respuesta del investigador a la incapacidad de los ordenadores de realizar la operación de integración. O dicho de forma más precisa, la simulación proporciona una aproximación numérica a las integrales, existiendo diferentes métodos que ofrecen diferentes propiedades, siendo aplicable cada uno de ellos a diferentes tipos de integrandos.

La simulación permite la estimación de modelos intratables por otras vías. Prácticamente cualquier modelo puede ser estimado mediante alguna forma de simulación. El investigador se ve liberado de esta forma de antiguas restricciones sobre la especificación del modelo, restricciones que reflejaban más la conveniencia matemática que la realidad económica de la situación estudiada. Esta nueva flexibilidad es un tremendo impulso para la investigación. Hace posible una representación más realista de la enorme

variedad de situaciones relativas a la elección que aparecen en el mundo. Permite al investigador obtener más información a partir de un conjunto de datos y, en muchos casos, permite afrontar problemas hasta ahora inabordables. Esta flexibilidad supone, sin embargo, una nueva carga para el investigador. En primer lugar, los nuevos métodos son en sí mismos más complicados que los anteriores, y utilizan numerosos conceptos y procedimientos que no se estudian en cursos de econometría típicos. Entender las diferentes técnicas – sus ventajas y limitaciones, y las relaciones entre ellas – es importante para escoger el método apropiado para un caso práctico específico y para desarrollar nuevos métodos cuando ninguno de los existentes parece apropiado. El propósito de este libro es ayudar al lector a lo largo de este camino.

En segundo lugar, para implementar un nuevo método o una variante de un método existente, el investigador necesita ser capaz de programar el procedimiento mediante software. Esto significa que el investigador a menudo necesitará conocer cómo funciona desde un punto de vista computacional la estimación mediante máxima verosimilitud (*maximum likelihood*) y otros métodos de estimación, cómo programar modelos específicos y cómo modificar programas existentes para representar variaciones en el comportamiento. Algunos modelos, como por ejemplo el logit mixto o el probit puro (adicionalmente al logit estándar), están implementados en paquetes de software estadístico disponibles comercialmente. De hecho, el código de estos y otros modelos, así como manuales y datos de ejemplo, están disponibles (de forma gratuita) en mi página web <http://elsa.berkeley.edu/~train>. Cuando sea apropiado, los investigadores deberían usar código ya disponible en lugar de escribir su propio código. Sin embargo, el valor real del nuevo enfoque dado a los modelos de elección es la capacidad de crear modelos a medida. Las tareas de cálculo y programación que se necesitan para implementar un nuevo modelo no son difíciles por norma general. Un objetivo importante del libro es enseñar estas capacidades como parte integral de la explicación de los propios modelos. Personalmente, considero que programar es extremadamente valioso a nivel pedagógico. El proceso de programación de un modelo me ayuda a comprender cómo funciona exactamente, las motivaciones e implicaciones de su estructura, qué características constituyen los elementos esenciales que no pueden ser cambiados para preservar el enfoque básico, y qué características son arbitrarias y pueden ser fácilmente modificadas. Imagino que otras personas también aprenden de esta misma manera.

1.2 Probabilidades de elección e integración

Para centrar ideas, voy a establecer la base conceptual de los modelos de elección discreta y a mostrar cómo la integración entra en juego. Un agente (por ejemplo, una persona, una empresa, un decisor) afronta la necesidad de realizar una elección, o una serie de elecciones a lo largo del tiempo, entre varias opciones disponibles. Por ejemplo, un consumidor elige qué producto comprar entre varios disponibles; una empresa decide qué tecnología usar en su producción; un estudiante elige qué respuesta dar a un test de respuesta múltiple; un participante en una encuesta elige un número entero entre 1 y 5 en una pregunta con una escala tipo *likert*; un trabajador elige si debe continuar trabajando cada año o retirarse. Nos referiremos al resultado de la decisión o decisiones tomadas en cualquier situación de elección como y , indicando la opción elegida o la secuencia de opciones. Asumimos para los propósitos de este libro que la variable resultado es discreta en el sentido de que puede tomar un conjunto numerable de valores. Muchos de los conceptos que describimos son fácilmente generalizables a situaciones en las que la variable resultado es continua. Sin embargo, la notación y la terminología son diferentes cuando tratamos con variables continuas en lugar de discretas. Asimismo, las elecciones discretas generalmente revelan menos información sobre el proceso de elección que las elecciones con resultado continuo, por lo que habitualmente la econometría de la elección discreta es más compleja.

Nuestro objetivo es entender el proceso de comportamiento que conduce a la elección realizada por el agente. Tomamos para ello una perspectiva causal. Hay factores que colectivamente determinan, o causan, la elección del agente. Algunos de estos factores son observados por el investigador y otros no.

A los factores observados los llamaremos x , y a los factores no observados ε . Los factores se relacionen con la elección del agente a través de una función $y = h(x, \varepsilon)$. Esta función la denominaremos proceso de comportamiento (*behavioral process*). Es determinista en el sentido de que dado x y ε , la elección del agente está totalmente determinada.

Pero dado que ε no ha sido observado, la elección del agente no es determinista y no puede ser predicha exactamente. En su lugar, calculamos la *probabilidad* de cualquier posible resultado. Los términos no observados son considerados aleatorios con una densidad de probabilidad $f(\varepsilon)$. La probabilidad de que el agente elija un resultado particular entre el conjunto de todos los posibles resultados es simplemente la probabilidad de que los factores no observados sean tales que hagan que el proceso de comportamiento arroje un resultado concreto: $P(y|x) = \text{Prob}(\varepsilon \text{ s.t. } h(x, \varepsilon) = y)$.

Podemos expresar esta probabilidad de una forma más práctica. Definamos una función indicadora $I[h(x, \varepsilon) = y]$ que toma el valor 1 cuando la expresión entre corchetes es verdadera y 0 cuando es falsa. Es decir, $I[\cdot] = 1$ si el valor de ε , combinado con x , induce al agente a elegir un resultado y , y $I[\cdot] = 0$ si el valor de ε , combinado con x , induce al agente a elegir otro resultado. De esta forma, la probabilidad de que el agente escoja el resultado y es simplemente el valor esperado de esta función indicadora, donde la esperanza se calcula respecto a todos los posibles valores de los factores no observados:

$$\begin{aligned} P(y | x) &= \text{Prob}(I[h(x, \varepsilon) = y] = 1) \\ (I.1) \quad &= \int I[h(x, \varepsilon) = y] f(\varepsilon) d\varepsilon \end{aligned}$$

Expresada de esta forma, la probabilidad es una integral, concretamente una integral de un indicador del resultado del proceso de comportamiento sobre todos los posibles valores de los factores no observados.

Para calcular esta probabilidad, debemos evaluar esta integral. Existen tres posibilidades para hacerlo.

1.2.1 Cálculo basado completamente en una expresión cerrada

Para ciertas especificaciones de h y f , la integral puede expresarse de forma cerrada. En esos casos, la probabilidad de elección puede calcularse de forma exacta a partir de dicha fórmula. Por ejemplo, consideremos un modelo logit binario relativo a si una persona realiza una acción o no, por ejemplo comprar un nuevo producto. El modelo de comportamiento se especifica de la siguiente manera. La persona obtendría cierto beneficio neto, o utilidad, en caso de realizar la acción. Esta utilidad, que puede ser positiva o negativa, está constituida por una parte que es observada por el investigador, $\beta'x$, donde x es un vector de variables y β es un vector de parámetros, y una parte que no es observada, ε : $U = \beta'x + \varepsilon$. La persona realiza la acción sólo si la utilidad es positiva, es decir, sólo si emprender la acción le proporciona un beneficio neto. La probabilidad de que la persona realice la acción, dada la información que el investigador puede observar, es por lo tanto $P = \int I[\beta'x + \varepsilon > 0] f(\varepsilon) d\varepsilon$, donde f es la densidad de probabilidad de ε . Asumamos que ε se distribuye logísticamente, de manera que su densidad es $f(\varepsilon) = e^{-\varepsilon}/(1 + e^{-\varepsilon})^2$ con una distribución de probabilidad acumulada $F(\varepsilon) = 1/(1 + e^{\varepsilon})$. En este caso, la probabilidad de que la persona realice la acción será:

$$\begin{aligned} P &= \int I[\beta'x + \varepsilon > 0] f(\varepsilon) d\varepsilon \\ &= \int I[\varepsilon > -\beta'x] f(\varepsilon) d\varepsilon \end{aligned}$$

$$\begin{aligned}
 &= \int_{\varepsilon = -\beta'x}^{\infty} f(\varepsilon) d\varepsilon \\
 &= 1 - F(-\beta'x) = 1 - \frac{1}{1 + e^{\beta'x}} \\
 &= \frac{e^{\beta'x}}{1 + e^{\beta'x}}
 \end{aligned}$$

Para cualquier x , la probabilidad puede calcularse de forma exacta como $P = \exp(\beta'x)/(1 + \exp(\beta'x))$.

Otros modelos también tienen una expresión cerrada para las probabilidades. Los modelos logit multinomial (capítulo 3), logit jerárquico (capítulo 4) y logit ordenado (capítulo 7) son ejemplos destacados. Los métodos que describí en mi primer libro y que fueron la base del interés inicial que despertó el análisis de la elección discreta, se apoyaban casi exclusivamente en modelos con expresión cerrada para las probabilidades de elección. En general, sin embargo, la integral necesaria para el cálculo de probabilidades no puede ser expresada de forma cerrada. O siendo más precisos, debemos aplicar restricciones sobre el modelo de comportamiento h y la distribución de probabilidad de los términos aleatorios f para lograr que la integral tenga una expresión cerrada. Estas restricciones pueden hacer los modelos poco realistas en muchas situaciones.

1.2.2 Cálculo basado completamente en la simulación

En lugar de resolver la integral de forma analítica, es posible aproximar su resultado mediante simulación. La simulación es aplicable de una manera u otra a prácticamente cualquier especificación de h y f . La simulación se fundamenta en el hecho de que integrar sobre una densidad de probabilidad es una forma de promediar. Consideremos la integral $\bar{t} = \int t(\varepsilon)f(\varepsilon)d\varepsilon$, donde $t(\varepsilon)$ es un estadístico basado en ε con densidad de probabilidad $f(\varepsilon)$. Esta integral corresponde al valor esperado de t sobre todos los posibles valores de ε . Este promedio puede aproximarse de una forma intuitivamente directa. Tomemos múltiples realizaciones (valores al azar) de la variable aleatoria ε a partir de su distribución de probabilidad f , calculemos $t(\varepsilon)$ para cada valor, y promediamos los resultados. Este promedio simulado es un estimador no sesgado del promedio real. Este procedimiento aproxima el valor del promedio real a medida que se utilizan más y más valores en la simulación.

Este concepto de simulación de un promedio es la base de todos los métodos de simulación, por lo menos de todos los que consideramos en este libro. Tal y como se indica en la ecuación (1.1), la probabilidad de que se produzca un resultado concreto es un promedio del indicador $I[\cdot]$ sobre todos los posibles valores de ε . La probabilidad, cuando se expresa de esta forma, puede ser simulada directamente como sigue:

1. Extraemos un valor al azar de ε a partir de $f(\varepsilon)$. Etiquetamos este valor como ε^1 , donde el superíndice indica que es la primera realización.
2. Determinamos si $h(x, \varepsilon^1) = y$ usando este valor de ε . Si es así, creamos $I^1 = 1$; en caso contrario fijamos $I^1 = 0$.
3. Repetimos los pasos 1 y 2 muchas veces, hasta un total de R valores. El indicador obtenido para cada realización se etiqueta como I^r donde $r = 1, \dots, R$.

4. Calculamos el promedio de los I^r . Este promedio es la probabilidad simulada: $\check{P}(y|x) = \frac{1}{R} \sum_{r=1}^R I^r$. Es la proporción de veces que los valores extraídos al azar de los factores no observados, en combinación con las variables observadas x , han producido un resultado y .

Como veremos en los siguientes capítulos, este simulador, aunque es fácil de comprender, tiene algunas propiedades desafortunadas. Las probabilidades de elección a menudo pueden expresarse como promedios de otros estadísticos, en lugar de promedios de una función indicadora. Los simuladores basados en estos otros estadísticos se calculan de forma análoga, mediante la extracción de valores al azar de la densidad de probabilidad, calculando el estadístico, y promediando los resultados. El modelo probit (capítulo 5) es el ejemplo más representativo de un modelo estimado completamente por simulación. Varios métodos para simular las probabilidades del modelo probit han sido desarrollados basándose en promedios de varios estadísticos sobre varias densidades (relacionadas).

1.2.3 Cálculo basado parcialmente en la simulación, parcialmente en una expresión cerrada

Hasta ahora hemos presentado los dos polos opuestos: o resolvemos la integral analíticamente o mediante simulación. En muchas ocasiones, es posible hacer un poco de ambas cosas.

Supongamos que los términos aleatorios pueden descomponerse de dos partes, que llamaremos ε_1 y ε_2 . La densidad de probabilidad conjunta de estos dos términos sería $f(\varepsilon) = f(\varepsilon_1, \varepsilon_2)$. La densidad conjunta puede expresarse como el producto de una densidad marginal y una densidad condicionada: $f(\varepsilon_1, \varepsilon_2) = f(\varepsilon_1|\varepsilon_2) \cdot f(\varepsilon_1)$. Usando esta descomposición, la probabilidad de la ecuación (1.1) puede expresarse como

$$P(y|x) = \int I[h(x, \varepsilon) = y]f(\varepsilon)d\varepsilon$$

$$\int_{\varepsilon_1} \left[\int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y]f(\varepsilon_2|\varepsilon_1)d\varepsilon_2 \right] f(\varepsilon_1)d\varepsilon_1$$

Ahora supongamos que existe una expresión cerrada para la integral que se encuentra dentro de los corchetes grandes. Denominemos esta fórmula como $g(\varepsilon_1) \equiv \int_{\varepsilon_2} I[h(x, \varepsilon_1, \varepsilon_2) = y]f(\varepsilon_2|\varepsilon_1)d\varepsilon_2$, fórmula que está condicionada respecto al valor de ε_1 . La probabilidad se puede simular extrayendo valores al azar de $f(\varepsilon_1)$, calculando $g(\varepsilon_1)$ para cada realización, y promediando posteriormente los resultados.

Este procedimiento se denomina *partición conveniente del error* (*convenient error partitioning*, Train, 1995). La integral respecto a ε_2 dado ε_1 se calcula exactamente, mientras que la integral respecto a ε_1 se calcula mediante simulación. Esta aproximación al problema presenta ventajas claras respecto a la simulación completa. Las integrales analíticas son más precisas y más fáciles de calcular que las integrales simuladas. Es útil por lo tanto, cuando es posible, descomponer los términos aleatorios de manera que una parte de ellos pueda ser integrada analíticamente, aun cuando el resto de términos deban ser simulados. Logit mixto (capítulo 6) es un ejemplo representativo de modelo que usa esta descomposición de forma efectiva. Otros ejemplos son el probit binario sobre datos de un panel, a cargo de Gourieroux and Monfort (1993), y el análisis de respuestas ordenadas de Bhat (1999).

1.3 Esquema del libro

El análisis de elecciones discretas consta de dos tareas interrelacionadas: la especificación del modelo de comportamiento y la estimación de los parámetros del modelo. La simulación juega un papel en ambas tareas. Por una parte, la simulación permite al investigador aproximar las probabilidades de

elección que surgen del modelo de comportamiento. Tal y como hemos mostrado, la capacidad de usar simulación da libertad al investigador para especificar modelos sin la restricción de tener que trabajar con probabilidades que tengan necesariamente una expresión cerrada. Por otra parte, la simulación también entra en juego en la tarea de estimación. Las propiedades de un estimador, como por ejemplo el estimador de máxima verosimilitud, pueden cambiar cuando se utilizan probabilidades simuladas en lugar de las probabilidades reales. Comprender estos cambios y mitigar los efectos negativos, es importante para el investigador. En algunos casos, como en los procedimientos Bayesianos, el estimador mismo es una integral sobre una densidad (en contraposición a los casos en los que la probabilidad de elección es una integral). La simulación permite implementar estos estimadores incluso cuando la integral que define el estimador no tiene una expresión cerrada.

Este libro se organiza en torno a estas dos tareas. La Parte I describe modelos de comportamiento que han sido propuestos para describir el proceso de elección. Los capítulos en esta sección van desde el modelo más simple, logit, hasta modelos progresivamente más generales y consecuentemente más complejos. Dedicamos un capítulo a cada uno de los siguientes modelos: logit, la familia de modelos generalizados de valor extremo (cuyo miembro más destacado es el logit jerárquico), probit y logit mixto. Esta parte del libro finaliza con un capítulo titulado “Variaciones sobre el tema”, que cubre una variedad de modelos que se construyen sobre los conceptos explicados en los capítulos precedentes. El objetivo de este capítulo va más allá de simplemente introducir varios modelos nuevos. El capítulo ilustra el concepto subyacente en todo el libro, a saber, que los investigadores necesitan no confiar únicamente en las pocas especificaciones comúnmente disponibles en software comercial, sino que pueden diseñar modelos que reflejen la singularidad de la configuración, los datos y los objetivos de su proyecto, escribiendo su propio código y usando simulación cuando se requiera.

La Parte II describe la estimación de los modelos de comportamiento. En primer lugar se aborda la maximización numérica, dado que la mayor parte de procedimientos de estimación implican la maximización de alguna función, como por ejemplo la función logaritmo de la verosimilitud (*log-likelihood*). A continuación describimos procedimientos para extraer valores al azar de diferentes tipos de densidades de probabilidad, lo cual es la base de la simulación. Este capítulo también describe diferentes tipos de extracciones de valores al azar, incluyendo variantes del método de antitéticos y las secuencias cuasi-aleatorias, que nos proporcionan mayor precisión en la simulación que el uso de valores aleatorios independientes. A continuación abordamos la estimación asistida por simulación, estudiando en primer lugar los procedimientos clásicos, incluyendo la máxima verosimilitud simulada, el método de momentos simulados y el método de puntuaciones simuladas, y posteriormente los procedimientos Bayesianos, incluyendo los métodos de Monte Carlo – Cadena de Markov. Hasta este punto del libro, asumimos que las variables explicativas son exógenas, es decir, independientes de factores no observados. El capítulo 13, que es nuevo en esta segunda edición, examina la endogeneidad, identificando situaciones en las que los factores no observados están correlacionados con las variables explicativas y describiendo métodos de estimación apropiados para estas situaciones, incluyendo el enfoque BLP, las funciones de control y la máxima verosimilitud con información completa. El capítulo final, que también es nuevo, muestra cómo los algoritmos EM, usados extensamente en otras áreas de la estadística, pueden ser de ayuda para modelos de elección complejos, incluyendo la estimación no paramétrica de la distribución de preferencias entre agentes. La simplicidad y la potencia de los algoritmos EM al ser aplicados a modelos de elección hacen de este capítulo un final apropiados para el libro.

1.4 Un par de notas

A lo largo de todo el libro, me refiero al investigador como “ella” y al decisor como “él”. Este uso, además de ser comparativamente neutral en relación al género (o al menos simétricamente no inclusivo), permite referirnos a ambos sujetos en la misma frase sin confusión.

Muchos colegas han proporcionado comentarios y sugerencias valiosas para este libro. Estoy muy agradecido por su ayuda. Gracias a Greg Allenby, Moshe Ben-Akiva, Chandra Bhat, Denis Bolduc, David Brownstone, Siddhartha Chib, Jon Eisen-Hecht, Florian Heiss, Stephane Hess, David Hensher, Joe Herriges, Rich Johnson, Frank Koppelman, Jordan Louviere, Aviv Nevo, Juan de Dios Ortúzar, John Rose, Ric Scarpa, Ken Small, Joan Walker, Cliff Winston, Joachim Winter y a los estudiantes de mi curso de econometría.