6

Logit mixto

6.1 Probabilidades de elección

Logit mixto es un modelo muy flexible que puede aproximar cualquier modelo de utilidad aleatoria (McFadden y Train, 2000). Este modelo elude las tres limitaciones del modelo logit estándar, permitiendo variación aleatoria de preferencias, patrones de sustitución no restringidos y correlación entre factores no observados a lo largo del tiempo. A diferencia de probit, no está limitado a distribuciones normales. Su formulación es sencilla y la simulación de sus probabilidades de elección es computacionalmente simple.

Al igual que probit, el modelo logit mixto se conoce desde hace muchos años, pero sólo se ha convertido en un modelo plenamente aplicable con la llegada de la simulación. Según parece, las primeras aplicaciones reales del logit mixto fueron los modelos de demanda de automóviles creados conjuntamente por Boyd y Mellman (1980) y Cardell y Dunbar (1980). En estos estudios, las variables explicativas no variaban entre decisores y la variable dependiente observada era la cuota de mercado y no las elecciones individuales de los clientes. Como resultado, la integración computacionalmente intensiva inherente al modelo logit mixto (como se explica más adelante) sólo fue necesario llevarla a cabo una única vez para el mercado como un todo, en lugar de realizarla para cada decisor de la muestra. Las primeras aplicaciones sobre datos a nivel de consumidor individual, como Train et al. (1987a) y Ben-Akiva et al. (1993), incluían sólo una o dos dimensiones de integración, las cuales podían calcularse por cuadratura numérica. Las mejoras en la velocidad de las computadoras y en el conocimiento de los métodos de simulación han permitido poder utilizar toda la potencia del modelo logit mixto. Entre los estudios que evidencian esta potencia están los de Bhat (1998a) y Brownstone y Train (1999) sobre datos transversales (elecciones en un único período de tiempo), y Erdem (1996), Revelt & Train (1998) y Bhat (2000), sobre datos de panel. La descripción que se ofrece en el presente capítulo se basa en gran medida en Train (1999).

Los modelos logit mixtos pueden formularse bajo diversas especificaciones de comportamiento y cada formulación proporciona una interpretación particular. El modelo logit mixto se *define* sobre la base de la forma funcional de sus probabilidades de elección. Cualquier especificación de comportamiento cuya formulación de las probabilidades de elección adopte esta forma particular se denomina un modelo logit mixto.

Las probabilidades del logit mixto son las integrales de las probabilidades logit estándar sobre una densidad de probabilidad de los parámetros. Dicho de manera más explícita, un modelo logit mixto es cualquier modelo cuyas probabilidades de elección se puedan expresar en la forma

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta,$$

donde $L_{ni}(\beta)$ es la probabilidad logit evaluada en los parámetros ß:

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}}$$

y $f(\beta)$ es una función de densidad de probabilidad. $V_{ni}(\beta)$ es la parte observada de la utilidad, que depende de los parámetros β. Si la utilidad es lineal en β, entonces $V_{ni}(\beta) = \beta x_{ni}$. En este caso, la probabilidad del modelo logit mixto toma su forma habitual:

(6.1)
$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} \right) f(\beta) d\beta,$$

La probabilidad del modelo logit mixto es un promedio ponderado de la fórmula logit evaluada en diferentes valores de β, con los pesos dados por la densidad $f(\beta)$. En la literatura estadística, la media ponderada de varias funciones se llama una función mixta (*mixed function* o *mixture function*) y la densidad que proporciona los pesos se llama la distribución de mezcla o mixtura (*mixing distribution*). El modelo logit mixto es una mezcla de la función logit evaluada en diferentes βs con $f(\beta)$ como distribución de mezcla.

Logit estándar es un caso especial de logit mixto en el que la distribución de mezcla $f(\beta)$ degenera en unos parámetros fijos b: $f(\beta) = 1$ para $\beta = b$ y 0 para $\beta \neq b$. La probabilidad elección (6.1) se convierte en la fórmula logit simple

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}}$$

La distribución de mezcla $f(\beta)$ puede ser discreta, con β tomando un conjunto finito de posibles valores. Supongamos que β toma M valores posibles etiquetados como $b_1, \dots, b_M$ con una probabilidad s_m de que $\beta = b_m$. En este caso, el modelo logit mixto se convierte en el *modelo de clases latentes* (*latent class model*), que ha sido popular durante mucho tiempo en psicología y marketing; algunos ejemplos los proporcionan Kamakura y Russell (1989) y Chintagunta et al. (1991). La probabilidad de elección en este caso resulta

$$P_{ni} = \sum_{m=1}^{M} s_m \left(\frac{e^{b_m' x_{ni}}}{\sum_{j=1}^{J} e^{b_m' x_{nj}}} \right).$$

Esta especificación resulta útil si hay M segmentos de población, cada uno de los cuales tiene su propio comportamiento de elección o preferencias. La proporción de la población que pertenece al segmento m es s_m, proporción que el investigador puede estimar en el modelo junto con las bs para cada segmento.

En la mayoría de aplicaciones que realmente han sido denominadas logits mixtos (tales como las citadas en los párrafos introductorios de este capítulo), se especifica que $f(\beta)$ sea continua. Por ejemplo, la densidad de β se puede especificar que sea una distribución normal con media b y covarianza W. La probabilidad de elección bajo esta densidad se convierte en

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}} \right) \phi(\beta | b, W) d\beta,$$

donde $\phi(\beta|b, W)$ es la densidad normal con media b y covarianza W.El investigador estima b y W. Distribuciones como la log-normal, la uniforme, la triangular, la gamma o cualquier otra pueden ser utilizadas del mismo modo. Como se verá en la sección 6.5, especificando apropiadamente las variables explicativas y la densidad, el investigador puede representar a través de un modelo logit mixto cualquier comportamiento orientado a maximizar la utilidad, así como muchas formas de comportamiento no relacionadas con la maximización de la utilidad.

McFadden y Train (2000) y Chesher & Santos - Silva (2002) han desarrollado varios test que muestran la necesidad de usar una distribución de mezcla no degenerada en parámetros fijos, así como la adecuación de cualquier distribución específica dada. Asimismo, varios estudios han comparado las distribuciones de mezcla discretas y continuas en el contexto de modelos logit mixtos; véase, por ejemplo, Wedel y Kamakura (2000) y Andrews et al. (2002).

Una cuestión terminológica surge en relación a los modelos logit mixtos. Hay dos conjuntos de parámetros en un modelo logit mixto. En primer lugar, tenemos los parámetros β, que entran en la fórmula logit. Estos parámetros tienen densidad $f(\beta)$. El segundo conjunto son parámetros que describen esta densidad. Por ejemplo, si β se distribuye normalmente con media b y covarianza W, entonces b y W son parámetros que describen la densidad $f(\beta)$. Por lo general (aunque no siempre, como señalaremos a continuación) el investigador está interesado en la estimación de los parámetros de f.

Denotemos los parámetros que describen la densidad de β como θ. La forma más adecuada para referirse a esta densidad es $f(\beta|\theta)$. Las probabilidades de elección del modelo logit mixto no dependen de los valores de β. Estas probabilidades son $P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta$, que son funciones de θ. Los parámetros β son las variables de integración y desaparecen del resultado. Por lo tanto, las βs son similares a la ε_{nj}s, en el sentido en que ambos son términos aleatorios que se integran para obtener la probabilidad de elección.

Bajo ciertas formulaciones del modelo logit mixto, los valores de β tienen un significado interpretable como una representación de las preferencias individuales de los decisores. En estos casos, el investigador desearía obtener información acerca de las βs para cada decisor de la muestra, así como los parámetros θ que describen la distribución de las βs entre decisores. En el capítulo 11, se describe la forma en que el investigador puede obtener esta información a partir de estimaciones de θ y de las elecciones observadas de cada decisor. En el presente capítulo se describe la estimación e interpretación de θ, usando procedimientos clásicos de estimación. En el capítulo 12 se describen los procedimientos bayesianos que proporcionan información sobre θ y sobre la β de cada decisor simultáneamente.

6.2 Coeficientes Aleatorios

La probabilidad del modelo logit mixto puede obtenerse bajo la hipótesis de un comportamiento orientado a la maximización de la utilidad, de varias formas que son formalmente equivalentes pero que ofrecen diferentes interpretaciones. La formulación más directa y más ampliamente utilizada en estudios recientes, se basa en coeficientes aleatorios. El decisor se enfrenta a una elección entre J alternativas. La utilidad que obtiene la persona n de la alternativa j se especifica como

$$U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj},$$

donde x_{nj} son variables observadas que se relacionan con la alternativa y el decisor, β_n es un vector de coeficientes de estas variables para la persona n que representa las preferencias de esa persona y ε_{nj} es un término aleatorio de tipo valor extremo iid. Los coeficientes varían entre decisores de la población con densidad $f(\beta)$. Esta densidad es una función de los parámetros θ que representan, por ejemplo, la media y la covarianza de las βs en la población. Esta especificación es igual a la del logit estándar, excepto que las βs varían entre decisores en lugar de ser fijas.

El decisor conoce el valor de su propia β_n y de las ε_{nj}s para toda alternativa j, y elige la alternativa i si y sólo si $U_{ni} > U_{nj}\ \forall j \neq i$. El investigador observa las x_{nj} pero no las β_ns o los ε_{nj}s. Si el investigador observase las β_ns, entonces la probabilidad de elección sería logit estándar, ya que los ε_{nj}s son valor extremo iid. Es decir, la probabilidad *condicionada* sobre β_n es

$$L_{ni}(\beta_n) = \frac{e^{\beta_n' x_{ni}}}{\sum_{j=1}^{J} e^{\beta_n' x_{nj}}}$$

Sin embargo, el investigador no conoce las β_ns y por lo tanto no puede condicionar sobre β. Por eso, la probabilidad de elección no condicionada es la integral de $L_{ni}(\beta_n)$ sobre todos los posibles valores de β_n:

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta) d\beta,$$

que es la probabilidad del modelo logit mixto (6.1).

El investigador especifica una distribución para los coeficientes y estima los parámetros de esa distribución. En la mayoría de los usos de este modelo, como Revelt & Train (1998), Mehndiratta (1996) y Ben-Akiva y Bolduc (1996), $f(\beta)$ se ha especificado como normal o log-normal : $\beta \sim N(b, W)$ o $\ln \beta \sim N(b, W)$ con parámetros b y W estimados. La distribución logarítmica normal es útil cuando se sabe que el coeficiente tiene el mismo signo para todos los decisores, como por ejemplo un coeficiente de precio que se sabe que es negativo para todo el mundo. Revelt & Train (2000), Hensher & Greene (2003), y Train (2001) han utilizado distribuciones triangulares y uniformes. Con la densidad uniforme, β se distribuye uniformemente entre $b - s$ y $b + s$, donde la media b y la extensión s deben ser estimadas. La distribución triangular tiene una densidad positiva que comienza en $b - s$, aumenta linealmente hasta b y luego desciende linealmente hasta $b + s$, tomando la forma de una tienda de campaña o triángulo. La media b y la extensión s se estiman, como en el caso de la distribución uniforme, pero la densidad es puntiaguda en lugar de plana. Estas densidades tienen la ventaja de estar limitadas por ambos lados, lo que evita el problema que puede surgir con las distribuciones normales y log-normales, las cuales pueden generar coeficientes anormalmente grandes para algún grupo de decisores. Al limitar s = b, el investigador puede asegurar que los coeficientes tienen el mismo signo para todos los decisores. Siikamaki (2001) y Siikamaki y Layton (2001) utilizan la distribución de Rayleigh (Johnson et al., 1994), que se encuentra en un solo lado del cero como la distribución log-normal pero, tal y como hallaron estos investigadores, puede resultar más simple a efectos de estimación que la log-normal. Revelt (1999) utilizó distribuciones normales truncadas. Como indican estos ejemplos, el investigador es libre de especificar una distribución que satisfaga sus expectativas sobre el comportamiento de su propia situación de elección.

Las variaciones en las preferencias que están relacionadas con los atributos observados de los decisores se capturan a través de la especificación de variables explicativas y/o la distribución de mezcla. Por ejemplo, el costo puede dividirse por los ingresos del decisor para permitir que el valor o la importancia relativa de los costos disminuya a medida que aumenta el ingreso. El coeficiente de esta variable aleatoria pasa a representar la variación del valor que las personas con el mismo nivel de ingresos otorgan al precio. La valoración media del costo se reduce con el aumento del ingreso mientras que la varianza entorno a la media es fija. Los atributos observados del decisor también pueden entrar en $f(\beta)$, de manera que los momentos de orden superior de la variación de las preferencias también puedan depender de los atributos del decisor. Por ejemplo, Bhat (1998a, 2000) especifica una $f(\beta)$ log-normal con media y varianza en función de las características del decisor.

6.3 Componentes de error

Un modelo logit mixto se puede utilizar sin una interpretación de coeficientes aleatorios, simplemente como una representación de los componentes de error que crea correlaciones entre las utilidades de diferentes alternativas. La utilidad se especifica como

$$U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj},$$

donde x_{nj} y z_{nj} son vectores que contienen variables observadas en relación a la alternativa j, α es un vector de coeficientes fijos, μ es un vector de términos aleatorios con media cero y ε_{nj} es de tipo valor extremo iid. Los términos en z_{nj} son componentes de error que, junto con ε_{nj}, definen la parte estocástica de la utilidad. Es decir, la parte no observada (aleatoria) de utilidad es $\eta_{nj} = \mu_n' z_{nj} + \varepsilon_{nj}$, que puede estar correlacionada entre alternativas en función de la especificación de z_{nj}. Para el modelo logit estándar, z_{nj} es idénticamente igual a cero, de modo que no hay correlación en la utilidad entre alternativas. Esta falta de correlación da lugar a la propiedad IIA y sus patrones de sustitución restrictivos. Con componentes de error distintos de cero, la utilidad está correlacionada entre alternativas: $\mathrm{Cov}(\eta_{ni}, \eta_{nj}) = \mathrm{E}(\mu_n' z_{ni} + \varepsilon_{ni})(\mu_n' z_{nj} + \varepsilon_{nj}) = z_{ni} W z_{nj}$, donde W es la covarianza de μ_n. La utilidad está correlacionada entre alternativas incluso cuando los componentes de error son independientes (como sucede en la mayoría de las especificaciones), de forma que W es diagonal.

Varios patrones de correlación, y por lo tanto varios patrones de sustitución, se pueden obtener mediante la elección apropiada de las variables que entran como componentes de error. Por ejemplo, un modelo análogo al logit jerárquico se obtiene mediante la especificación de una variable indicadora (*dummy*) para cada nido que sea igual a 1 para cada alternativa en el nido y cero para las alternativas fuera del nido. Con K nidos no solapados, los componentes de error son $\mu_n' z_{nj} = \sum_{k=1}^{K} \mu_{nk} d_{jk}$, donde $d_{jk} = 1$ si j está en el nido k y cero en caso contrario. Es conveniente en esta situación especificar que los componentes de error se distribuyan como normales independientes: $\mu_{nk}\ iid\ N(0, \sigma_k)$. La cantidad aleatoria μ_{nk} entra en la utilidad de cada alternativa del nido k, induciendo correlación entre estas alternativas. No entra en ninguna de las alternativas de otros nidos, con lo cual no induce correlación entre alternativas del nido con alternativas fuera del nido. La varianza σ_k capta la magnitud de la correlación. Desempeña un papel análogo al coeficiente de valor inclusivo de los modelos logit jerárquicos.

Para ser más precisos, la covarianza entre dos alternativas en el nido k es $\mathrm{Cov}(\eta_{ni}, \eta_{nj}) = \mathrm{E}(\mu_k + \varepsilon_{ni})(\mu_k + \varepsilon_{nj}) = \sigma_k$. La varianza para cada una de las alternativas en el nido k es $\mathrm{Var}(\eta_{ni}) = \mathrm{E}(\mu_k + \varepsilon_{ni})^2 = \sigma_k + \pi^2/6$, ya que la varianza del término de valor extremo, ε_{ni}, es $\pi^2/6$ (véase la

Sección 3.1). La correlación entre dos alternativas cualesquiera dentro del nido de k, es por lo tanto $\sigma_k/(\sigma_k+\pi^2/6)$. Restringir la varianza de cada componente de error de cada nido para que sea igual en todos los nidos (es decir, forzar que $\sigma_k=\sigma, k=1,\dots,K$) es análogo a restringir el coeficiente de log-suma para que sea igual para todos los nidos en un modelo logit jerárquico. Esta restricción también asegura que el modelo logit mixto esté normalizado para la escala y el nivel.

Permitir diferentes varianzas de las cantidades aleatorias de nidos diferentes es análogo a permitir que el coeficiente de valor inclusivo difiera entre los nidos en un logit jerárquico. Un efecto análogo al de los nidos solapados se captura con variables indicadoras que identifiquen conjuntos solapados de alternativas, como en Bhat (1998a). Un modelo análogo al logit heterocedástico (visto en la sección 4.5) se obtiene mediante la introducción de un componente de error para cada alternativa. Walker et al. (2007) proporcionan orientación sobre cómo especificar estas variables de manera apropiada.

Las especificaciones basadas en componentes de error y en coeficientes aleatorios son formalmente equivalentes. Basándonos en coeficientes aleatorios, la utilidad se especifica como $U_{nj}=\beta_n' x_{nj}+\varepsilon_{nj}$ con β_n aleatorio. Los coeficientes β_n se pueden descomponer entre su media α y las desviaciones μ_n, de manera que $U_{nj}=\alpha_n' x_{nj}+\mu_n' x_{nj}+\varepsilon_{nj}$, que tiene componentes de error definidos por $z_{nj}=x_{nj}$. En el sentido contrario, basándonos en componentes de error, la utilidad es $U_{nj}=\alpha' x_{nj}+\mu_n' z_{nj}+\varepsilon_{nj}$, lo que equivale a un modelo de parámetros aleatorios con coeficientes fijos para las variables x_{nj} y coeficientes aleatorios con media cero para las variables z_{nj}. Si x_{nj} y z_{nj} se solapan (en el sentido de que algunas variables entran tanto en x_{nj} como en z_{nj}), los coeficientes de estas variables pueden considerarse que varían de forma aleatoria con media α y la misma distribución de μ_n alrededor de sus medias.

Aunque coeficientes aleatorios y componentes de error son formalmente equivalentes, la forma en que un investigador piensa en el modelo afecta a la especificación del logit mixto. Por ejemplo, cuando se piensa en términos de parámetros aleatorios, es natural permitir que el coeficiente de cada variable pueda variar e incluso permitir correlaciones entre coeficientes. Este es el enfoque adoptado por Revelt & Train (1998). Sin embargo, cuando el objetivo principal es representar patrones de sustitución apropiadamente a través de la utilización de componentes de error, el énfasis se pone en especificar variables que puedan inducir correlación entre alternativas de una manera lo más simple posible como para proporcionar patrones de sustitución suficientemente realistas. Este es el enfoque adoptado por Brownstone y Train (1999). Los objetivos eran diferentes en estos estudios: Revelt y Train estaban interesados en el patrón de preferencias, mientras que Brownstone y Train estaban más preocupados por la predicción. El número de variables explicativas también difirió en ambos casos, con Revelt y Train examinando 6 variables, lo que permitía que la estimación de la distribución conjunta de sus coeficientes fuese una meta razonable, mientras que Brownstone y Train incluían 26 variables en su análisis. Aspirar a estimar la distribución de 26 coeficientes no es razonable, y sin embargo, pensar en términos de parámetros aleatorios en lugar de componentes de error puede llevar al investigador a tales expectativas. Es importante recordar que la distribución de mezcla, ya sea motivada por parámetros aleatorios o por componentes de error, captura la varianza y las correlaciones de los factores no observados. Hay un límite natural en cuánto puede aprenderse sobre las cosas que no son observadas.

6.4 Patrones de sustitución

Logit mixto no presenta independencia de alternativas irrelevantes (IIA) o los patrones de sustitución restrictivos de logit. El ratio de las probabilidades de elección en los modelos logit mixtos, P_{ni}/P_{nj}, depende de todos los datos, incluidos los atributos de las alternativas que no son i o j. Los denominadores de la fórmula logit se encuentran dentro de las integrales, por lo que no se anulan. El porcentaje de cambio en la probabilidad de una de las alternativas dado un cambio porcentual en el atributo m-ésimo de otra alternativa, es

$$E_{nix_{nj}^m} = -\frac{x_{nj}^m}{P_{ni}} \int \beta^m L_{ni}(\beta) L_{nj}(\beta) f(\beta) d\beta,$$

$$= -x_{nj}^m \int \beta^m L_{nj}(\beta) \left[\frac{L_{ni}(\beta)}{P_{ni}}\right] f(\beta) d\beta,$$

donde β^m es el elemento m-ésimo de β. Esta elasticidad es diferente para cada alternativa i. Una reducción del diez por ciento en una alternativa no implica necesariamente (como en logit) una reducción del diez por ciento en cada una de las otras alternativas. En este caso, el patrón de sustitución depende de la especificación de las variables y de la distribución de mezcla, y ambas puede determinarse empíricamente.

Observe que el porcentaje de cambio en la probabilidad depende de la correlación entre $L_{ni}(\beta)$ y $L_{nj}(\beta)$ a través de diferentes valores de β, la cual está determinada por la especificación que el investigador hace de las variables y la distribución de mezcla. Por ejemplo, para representar una situación en la que una mejora en la alternativa j reduce proporcionalmente más la alternativa i que la alternativa k, el investigador puede especificar un elemento de x que correlacione positivamente entre i y j, pero que no correlacione o correlacione negativamente entre k y j, con una distribución de mezcla que permita variar al coeficiente de esta variable.

6.5 Aproximación de cualquier modelo de utilidad aleatoria

McFadden y Train (2000) muestran que cualquier modelo de utilidad aleatoria (*random utility model*, RUM) puede ser aproximado con cualquier grado de precisión por un modelo logit mixto, con la elección apropiada de las variables y la distribución de mezcla. Esta demostración es análoga a las aproximaciones consistentes con RUM proporcionadas por Dagsvik (1994). Es posible proporcionar una explicación intuitiva fácilmente. Supongamos que el verdadero modelo es $\mathrm{U_{nj}} = \alpha_n' \mathrm{z_{nj}}$, donde $\mathrm{z_{nj}}$ son variables relacionadas con la alternativa j y α sigue una distribución $f(\alpha)$ cualquiera. Cualquier RUM puede expresarse de esta forma. (La notación más tradicional $\mathrm{U_{nj}} = \beta_n' \mathrm{x_{nj}} + \varepsilon_{nj}$ se obtiene definiendo $\mathrm{z_{nj}'} = \langle \mathrm{x_{nj}'}, d_j \rangle$, $\alpha' = \langle \beta_n', \varepsilon_{nj} \rangle$ y $\mathrm{f}(\alpha)$ como la densidad conjunta de β_ny ε_{nj} $\forall j$). Condicionada a α, la elección de la persona está totalmente determinada, dado que $\mathrm{U_{nj}}$ se conocería entonces para cada j. Por tanto, la probabilidad condicionada es

$$q_{ni}(\alpha) = I\left(\alpha_n' z_{ni} > \alpha_n' z_{nj} \ \forall j \neq i\right),$$

donde $\mathrm{I}(\cdot)$ es la función indicadora 1-0 de si se produce el evento entre paréntesis. Esta probabilidad condicionada es determinista en el sentido de que la probabilidad o es cero o es uno: condicionada a todos los términos aleatorios desconocidos, la elección del decisor está completamente determinada. La probabilidad de elección no condicionada es la integral de $\mathrm{q_{ni}}(\alpha)$ sobre α:

$$Q_{ni} = \int I\left(\alpha_n' z_{ni} > \alpha_n' z_{nj} \ \forall j \neq i\right) f(\alpha) d\alpha,$$

Podemos aproximar esta probabilidad con un modelo logit mixto. Escalemos la utilidad por un factor λ, de modo que $\mathrm{U_{nj}^*} = (\alpha/\lambda)' \mathrm{z_{nj}}$. Este escalado no cambia el modelo, ya que el comportamiento no se ve afectado por la escala de la utilidad. A continuación, agregamos un término tipo valor extremo iid: ε_{nj}.

La adición del término valor extremo no cambia el modelo, ya que cambia la utilidad de cada alternativa. Lo agregamos porque al hacerlo resulta un logit mixto. Y, como veremos (este es el propósito de la demostración), añadir el término valor extremo es inocuo. La probabilidad del modelo logit mixto basado en esta utilidad es

$$P_{ni} = \int \left(\frac{e^{(\alpha/\lambda)' z_{ni}}}{\sum_j e^{(\alpha/\lambda)' z_{nj}}} \right) f(\alpha) d\alpha,$$

A medida que λ se aproxima a cero, los coeficientes α/λ en la fórmula logit se hacen mayores y P_{ni} se parece cada vez más a un indicador 1-0 de la alternativa con mayor utilidad. Es decir, la probabilidad del modelo logit mixto P_{ni} se aproxima a la verdadera probabilidad Q_{ni} a medida que λ se aproxima a cero. Escalando los coeficientes al alza suficientemente, el logit mixto basado en estos coeficientes escalados se aproxima arbitrariamente al modelo verdadero. Srinivasan y Mahmassani (2005) utilizan este concepto de aumentar la escala de los coeficientes para demostrar que un modelo logit mixto puede aproximar un modelo probit; el concepto aplica en general a la aproximación de cualquier RUM.

Recuerde que hemos añadido un término valor extremo iid a la verdadera utilidad de cada alternativa. Estos términos cambian el modelo, porque la alternativa con mayor utilidad antes de que los términos fuesen añadidos puede no tener la mayor utilidad después (ya que se añade una cantidad diferente a cada utilidad). Sin embargo, al aumentar la escala de utilidad suficientemente, podemos estar totalmente seguros de que la adición de los términos valor extremo no tiene ningún efecto. Consideremos un ejemplo con dos alternativas. Supongamos, utilizando el verdadero modelo con su escala original, que la utilidad de la alternativa 1 es 0.5 unidades mayor que la utilidad de la alternativa 2, de modo que la alternativa 1 es la elegida. Supongamos que añadimos un término valor extremo para cada alternativa. Hay una probabilidad considerable, dada la varianza de estos términos aleatorios, de que el valor obtenido para la alternativa 2 exceda el de la alternativa 1 por lo menos en media unidad, de manera que la alternativa 2 sea ahora la que obtiene mayor utilidad en lugar de la alternativa 1. Por lo tanto, la adición de los términos valor extremo cambia el modelo, ya que cambia la alternativa que tiene mayor utilidad. Supongamos, sin embargo, que podemos aumentar la escala de la utilidad original por un factor 10 (es decir, $\lambda = 0.10$). La utilidad de la alternativa 1 supera ahora la utilidad de la alternativa 2 de 5 unidades en lugar de 0.5 unidades. Es muy poco probable que la adición de términos valor extremo a estas utilidades invierta esta diferencia. Es decir, es muy poco probable, de hecho casi imposible, que el valor de $\varepsilon_{\mathrm{n2}}$ que se agrega a la utilidad de la alternativa 2 sea mayor en 5 unidades al término $\varepsilon_{\mathrm{n1}}$ que se agrega a la utilidad de la alternativa 1. Si el re-escalado en un factor 10 no es suficiente para asegurar que la adición del término valor extremo no tenga ningún efecto, entonces las utilidades originales podrían re-escalarse en un factor 100 o 1000. En algún momento, encontraremos una escala para la que la suma de los términos valor extremo no tenga ningún efecto. Dicho de manera sucinta, la adición de un término valor extremo a la verdadera utilidad, que convierte el modelo en un logit mixto, no cambia la utilidad de manera significativa cuando la escala de la utilidad es lo suficientemente grande. Un logit mixto puede aproximar cualquier RUM simplemente ampliando suficientemente la escala de la utilidad.

Esta demostración no pretende sugerir que aumentar la escala de la utilidad es la forma en que el investigador procederá realmente cuando especifique un logit mixto como una aproximación al verdadero modelo. Más bien, la demostración simplemente indica que si no se pueden encontrar otros medios para especificar un modelo logit mixto que aproxime el verdadero modelo, entonces este procedimiento de cambio de escala puede ser usado para lograr la aproximación. Por lo general, un logit mixto se puede especificar de manera que refleje adecuadamente el verdadero modelo sin necesidad de recurrir a una escala aumentada de utilidad. Por ejemplo, el verdadero modelo por lo general contendrá algún término iid que se agrega a la utilidad de cada alternativa. Suponiendo una distribución tipo valor extremo para este término tal vez está lo suficientemente cerca de la realidad como para ser empíricamente indistinguible de otros supuestos de distribución para el término iid. En este caso, la escala de la utilidad se determina naturalmente por la varianza de este término iid. La tarea del investigador es simplemente encontrar las variables y una distribución de mezcla que capten las otras partes de la utilidad, es decir, las partes que están correlacionadas entre alternativas o heterocedásticas.

6.6 Simulación

El modelo logit mixto se adapta bien a los métodos de simulación para la estimación. La utilidad es $\mathrm{U_{nj}} = \beta_n' x_{nj} + \varepsilon_{nj}$, donde los coeficientes β_n se distribuyen con densidad $f(\beta|\theta)$, donde θ se refiere colectivamente a los parámetros de esta distribución (tales como la media y la covarianza de β). El investigador especifica la forma funcional $f(\cdot)$ y desea estimar los parámetros θ. Las probabilidades de elección son

$$P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta,$$

donde

$$L_{ni}(\beta) = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}}.$$

Las probabilidades se aproximan mediante simulación para cualquier valor dado de θ: (1) Extraiga al azar un valor β de $f(\beta|\theta)$, y etiquételo como β^r, con el superíndice $r = 1$ en referencia al primer valor extraído. (2) Calcule la fórmula logit $L_{ni}(\beta^r)$ con este valor. (3) Repita los pasos 1 y 2 múltiples veces y promedie los resultados. Este promedio es la probabilidad simulada:

$$\check{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r)$$

donde R es el número de valores extraídos al azar usados en la simulación. $\check{P}_{ni}$ es un estimador no sesgado de P_{ni} por la forma en que se construye. Su varianza disminuye a medida que aumenta R. Es estrictamente positivo, de modo que $\ln \check{P}_{ni}$ está definido, lo que es útil para aproximar a continuación la función log-verosimilitud. $\check{P}_{ni}$ es suave (dos veces diferenciable) en los parámetros θ y en las variables x, lo que facilita la búsqueda numérica de la máxima verosimilitud y el cálculo de elasticidades. Y la suma de $\check{P}_{ni}$ para todas las alternativas es uno, algo útil para hacer pronósticos.

Las probabilidades simuladas se insertan en la función log -verosimilitud para calcular una log-verosimilitud simulada:

$$SLL = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \ln \check{P}_{nj},$$

donde $d_{nj} = 1$ si n eligió j y cero en caso contrario. El estimador de máxima verosimilitud simulada (MSLE) es el valor de θ que maximiza SLL. Las propiedades de este estimador se tratan en el Capítulo 10. Generalmente, se extraen al azar valores diferentes para cada observación. Este procedimiento mantiene la independencia entre decisores de las probabilidades simuladas que entran en la SLL. Lee

(1992) describe las propiedades del MSLE cuando se utilizan para todas las observaciones los mismos valores extraídos al azar.

La probabilidad simulada de un logit mixto puede relacionarse con métodos de simulación de tipo aceptación-rechazo (AR). La simulación AR se describe en la Sección 5.6 para los modelos probit, pero es aplicable de forma más general. Para cualquier modelo de utilidad aleatoria, el simulador AR se construye como sigue: (1) Se extrae al azar un valor de los términos aleatorios. (2) Se calcula la utilidad de cada alternativa a partir de ese valor y se identifica la alternativa con mayor utilidad. (3) Los pasos 1 y 2 se repiten múltiples veces. (4) La probabilidad de elección simulada para una alternativa concreta se calcula como la proporción de valores extraídos para los que esa alternativa ha sido la de mayor utilidad. El simulador AR es no sesgado por construcción. Sin embargo, no es estrictamente positivo para cualquier número finito de valores extraídos. Tampoco es una función suave, sino una función escalonada: constante dentro de los rangos de parámetros para los que la identidad de la alternativa con mayor utilidad no cambia para cualquier valor extraído, y con saltos donde los cambios en los parámetros cambian la identidad de la alternativa de mayor utilidad. Los métodos numéricos de maximización basados en el simulador AR se ven perjudicados por estas características. Para hacer frente a estos problemas numéricos, el simulador AR puede ser suavizado reemplazando la función indicadora 0-1 por la fórmula logit. Tal y como vimos en la Sección 5.6.2, el simulador AR suavizado-logit puede aproximar el simulador AR hasta un nivel de similitud arbitrario mediante un re-escalado de la utilidad apropiado.

El simulador logit mixto puede verse como un simulador AR suavizado-logit de cualquier RUM: se extraen valores al azar de los términos aleatorios, se calculan las utilidades para estos valores, las utilidades calculadas se insertan en la fórmula logit y se promedian los resultados. El teorema que afirma que un logit mixto puede aproximar cualquier modelo de utilidad aleatoria (sección 6.5) puede ser visto desde esta perspectiva. Sabemos por la Sección 5.6.2 que el simulador AR suavizado-logit puede aproximarse arbitrariamente al simulador AR de cualquier modelo, siempre y cuando se escale suficientemente la utilidad. Dado que el simulador logit mixto es equivalente a un simulador AR suavizado-logit, el modelo logit mixto simulado puede estar arbitrariamente cerca del simulador AR de cualquier modelo.

6.7 Datos de panel

La especificación del modelo logit mixto se puede generalizar fácilmente para permitir elecciones repetidas de cada decisor de la muestra. La especificación más simple trata los coeficientes que entran en la utilidad como parámetros que varían entre personas pero que son constantes en situaciones entre situaciones de elección de una misma persona. La utilidad de la alternativa j en una situación de elección t por parte de una persona n es $U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$, con ε_{njt} siendo valor extremo iid a lo largo del tiempo, de las personas y de las alternativas. Considere una secuencia de alternativas, una para cada período de tiempo, $\mathbf{i} = \{i_1, \dots, i_T\}$. Condicionada a β, la probabilidad de que la persona haga esta secuencia de elecciones es el producto de fórmulas logit:

$$(6.2) \qquad \boldsymbol{L}_{ni}(\beta) = \prod_{t=1}^{T} \left[\frac{e^{\beta_n' x_{ni_t t}}}{\sum_j e^{\beta_n' x_{njt}}} \right],$$

dado que los ε_{njt}s son independientes en el tiempo. La probabilidad no condicionada es la integral de este producto sobre todos los valores de β:

$$(6.3) \qquad P_{ni} = \int \boldsymbol{L}_{ni}(\beta) f(\beta) d\beta.$$

La única diferencia entre un logit mixto con elecciones repetidas y uno con una sola elección por decisor es que el integrando implica un producto de fórmulas logit, una para cada período de tiempo, en lugar de sólo una fórmula logit. La probabilidad se simula de manera similar a la probabilidad con un único período de elección. Se extrae al azar un valor β de su distribución de probabilidad. Se calcula la fórmula logit para cada período y se calcula el producto de estas fórmulas logit. Este proceso se repite para múltiples valores extraídos y se promedian los resultados.

Es posible agregar a la utilidad variables exógenas pasadas y futuras en un período determinado para representar una respuesta diferida o un comportamiento anticipatorio, como se describe en la sección 5.5 en relación a un modelo probit con datos de panel. Sin embargo, a diferencia de probit, en un modelo logit mixto es posible añadir variables dependientes diferidas sin cambiar el procedimiento de estimación. Si condicionamos a β_n, los únicos términos aleatorios que quedan en el logit mixto son los términos ε_{nj}s, que son independientes en el tiempo. Una variable dependiente diferida que entre en U_{njt} no está correlacionada con estos términos de error restantes para el período t, ya que estos términos son independientes a lo largo del tiempo. Las probabilidades condicionadas (condicionadas a β) son, por lo tanto, las mismas de la ecuación (6.2), pero con x incluyendo variables dependientes diferidas. La probabilidad no condicionada es la integral de esta probabilidad condicionada sobre todos los valores de ß, que es justamente la ecuación (6.3). En este sentido, logit mixto es un modelo más conveniente que probit para la representación de la dependencia del estado, ya que las variables dependientes diferidas se pueden añadir al logit mixto sin ajustar la fórmula de probabilidad o el método de simulación. Erdem (1996) y Johannesson y Lundin (2000) explotan esta ventaja para examinar la formación de hábitos y la búsqueda de la variedad dentro de un logit mixto que también captura la variación aleatoria de preferencias.

Si las elecciones y los datos no se observan desde el inicio del proceso (es decir, desde la primera situación de elección que la persona afronta), es necesario resolver la cuestión de las condiciones iniciales, al igual que con probit. El investigador debe representar de alguna manera la probabilidad de la primera elección observada, que depende de las elecciones previas no observadas. Heckman y Singer (1986) proporcionan formas de manejar este problema. Sin embargo, cuando el investigador observa el proceso de elección desde el principio, el problema de las condiciones iniciales no se plantea. En este caso, el uso de variables dependientes diferidas para capturar la inercia en la elección u otros tipos de dependencia del estado, es algo sencillo para el modelo logit mixto. Los datos de preferencia declarada (es decir, las respuestas a una serie de hipotéticas situaciones de elección planteadas a los participantes en una encuesta) son un ejemplo claro de datos en los que el investigador observa toda la secuencia de elecciones.

En la especificación desarrollada hasta el momento, así como en casi todas las aplicaciones prácticas, se asume que los coeficientes β_n son constantes entre diferentes situaciones de elección para un mismo decisor. Este supuesto es apropiado si las preferencias del decisor son estables a lo largo del período de tiempo que comprende las elecciones repetidas. Sin embargo, es posible especificar que los coeficientes asociados a cada persona puedan variar con el tiempo de diversas maneras. Por ejemplo, las preferencias de cada persona pueden estar correlacionadas en serie entre situaciones de elección, por lo que la utilidad sería

$$U_{njt} = \beta_{nt} x_{njt} + \varepsilon_{njt},$$

$$\beta_{nt} = b + \tilde{\beta}_{nt},$$

$$\tilde{\beta}_{nt} = \rho \tilde{\beta}_{nt-1} + \mu_{nt},$$

donde b es fijo y μ_{nt} es iid sobre n y t. La simulación de la probabilidad de que la secuencia de elecciones se produzca se realiza como sigue:

1. Extraiga un valor al azar para μ_{n1}^r para el período inicial y calcule la fórmula logit para este período utilizando $\beta_{n1}^r = b + \mu_{n1}^r$.
2. Extraiga un valor al azar para μ_{n2}^r para el segundo período, calcule $\beta_{n2}^r = b + \rho\mu_{n1}^r + \mu_{n2}^r$ y luego calcule la fórmula logit en base a esta β_{n2}^r.
3. Continúe para todos los períodos de tiempo T.
4. Tome el producto de los T logits.
5. Repita los pasos 1-4 para numerosas extracciones de valores.
6. Promedie los resultados.

La carga que colocamos en la tarea de simulación es mayor que con coeficientes constantes en el tiempo para cada persona, requiriendo la realización de extracciones T veces.

6.8 Estudio de un caso

A modo ilustrativo, considere un modelo logit mixto relativo a las elecciones que los pescadores realizan sobre los sitios a los que ir a pescar (Train, 1999). La especificación emplea el uso de coeficientes aleatorios. La utilidad es $U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$, con coeficientes β_n que varían entre pescadores pero no entre las decisiones (los viajes realizados) de cada pescador. La probabilidad de la secuencia de los sitios elegidos por cada pescador viene dada por la ecuación (6.3).

La muestra consta de 962 viajes a ríos de Montana realizados por un total de 258 pescadores durante el período comprendido entre julio de 1992 y agosto de 1993. Se definieron un total de 59 posibles sitios en los ríos, con base a la situación geográfica y a otros factores relevantes. Cada sitio contiene uno o más tramos de ríos definidos en el sistema de información fluvial de Montana. Las siguientes variables entran como elementos de x para cada sitio:

1. Disponibilidad de peces, medida en unidades de 100 peces por cada 1000 pies de río.
2. Valoración estética, medida en una escala de 0 a 3, siendo 3 la valoración más alta.
3. Costo del viaje: Costo de viajar desde la residencia del pescador hasta el sitio de pesca, incluyendo la variable de costo de conducir (combustible, mantenimiento, neumáticos, aceite) y el valor del tiempo consumido en la conducción (con el tiempo valorado como un tercio del salario del pescador).
4. Indicador de que la “Guía del pescador en Montana” menciona el sitio como sitio de pesca mayor.
5. Número de campings por bloque en el sitio, tal y como se define bloque en el “U.S. Geological Survey (USGS)”.
6. Número de zonas recreativas estatales por bloque USGS.
7. Número de especies de pesca restringida (especies restringidas) en el sitio.
8. Logaritmo del tamaño del sitio, en bloques USGS.

Lógicamente, los coeficientes de las variables 4-7 pueden tener cualquier signo; por ejemplo, a algunos pescadores les puede gustar tener campings cerca mientras que otros pueden preferir la privacidad que proporciona no tener campings cercanos. A cada uno de estos coeficientes se le asigna una distribución normal independiente con media y desviación estándar a estimar. Se espera que los coeficientes de costo del viaje, disponibilidad de peces y valoración estética del sitio tengan el mismo signo para todos los pescadores, difiriendo entre pescadores sólo en sus magnitudes. A estos coeficientes se asignan distribuciones log-normales independientes. Se estiman la media y la desviación estándar del logaritmo del coeficiente y, a partir de estas estimaciones, se calculan la media y la desviación estándar del propio coeficiente. Dado que la distribución log-normal se define en el rango positivo y se espera que el costo del viaje tenga un coeficiente negativo para todos los pescadores, introducimos en el modelo el costo del viaje con el signo invertido (negativa del costo). Se asume que el coeficiente del logaritmo del tamaño del sitio sea fijo. Esta variable permite contemplar el hecho de que la probabilidad de visitar un sitio más grande sea mayor que la de un sitio pequeño, en igualdad de todos los demás parámetros. Permitir que el coeficiente de esta variable variase entre personas, aunque sería posible, no sería particularmente significativo. Una versión del modelo con coeficientes correlacionados la proporciona Train (1998). El modelo de elección del sitio forma parte de un modelo global, dado por Desvousges et al. (1996), de la elección conjunta de la frecuencia de viaje y la elección del sitio.

La simulación se llevó a cabo utilizando mil extracciones de valores al azar para cada pescador de la muestra. Los resultados se facilitan en la tabla 6.1. La desviación estándar de cada coeficiente aleatorio es altamente significativa, lo que indica que esos coeficientes en efecto varían en la población.

Tabla 6.1. Modelo logit mixto de la elección de sitios de pesca en ríos

Variable	Parámetro	Valor	Error estándar
Disponibilidad de peces	Media del ln(coeficiente)	−2.876	0.6066
	Desv. est. del ln(coeficiente)	1.016	0.2469
Valoración estética	Media del ln(coeficiente)	−0.794	0.2287
	Desv. est. del ln(coeficiente)	0.849	0.1382
Costo total (negativo)	Media del ln(coeficiente)	−2.402	0.0631
	Desv. est. del ln(coeficiente)	0.801	0.0781
Listado en la guía como sitio mayor	Media del coeficiente	1018	0.2887
	Desv. est. del coeficiente	2.195	0.3518
Campings	Media del coeficiente	0.116	0.3233
	Desv. est. del coeficiente	1.655	0.4350
Áreas de acceso	Media del coeficiente	−0.950	0.3610
	Desv. est. del coeficiente	1.888	0.3511
Especies restringidas	Media del coeficiente	−0.499	0.1310
	Desv. est. del coeficiente	0.899	0.1640
Log(tamaño)	Media del coeficiente	0.984	0.1077
Índice del coeficiente de verosimilitud		0.5018	
SLL en convergencia		−1932.33	

Consideremos en primer lugar los coeficientes distribuidos normalmente. Las medias estimadas y las desviaciones estándar de estos coeficientes proporcionan información sobre la proporción de la población que valora positivamente un atributo del sitio y la que lo valora negativamente. La distribución del coeficiente relativo a la variable que indica si el sitio es mencionado como importante en la “Guía del pescador en Montana” obtiene una media estimada de 1.018 y una desviación estándar estimada de 2.195, de tal manera que el 68 por ciento de la distribución está por encima de cero y el 32 por ciento por debajo. Esto implica que estar catalogado como un sitio importante en la “Guía del pescador en Montana” es un incentivo positivo para cerca de dos tercios de los pescadores y un factor negativo para el otro tercio, que aparentemente prefiere la soledad. La presencia de campings es preferida por aproximadamente la mitad (53 por ciento) de los pescadores y es evitada por la otra mitad. Y se estima que cerca de un tercio de los pescadores (31 por ciento) prefieren tener numerosas áreas recreativas, mientras que las otras dos terceras partes prefieren tener menos áreas.

Consideremos ahora los coeficientes definidos con distribuciones log-normales. Un coeficiente β^k sigue una distribución log-normal si el logaritmo de β^k se distribuye normalmente. Parametrizamos la distribución log-normal en términos de la distribución normal subyacente. Es decir, estimamos los parámetros m y s que representan la media y la varianza del logaritmo del coeficiente: $\ln \beta^k \sim N(m, s)$. La media y la varianza de β^k se obtienen acto seguido a partir de las estimaciones de m y s. La mediana es $\exp(\mathrm{m})$, la media es $\exp(\mathrm{m}+\mathrm{s}/2)$ y la varianza es $\exp(2\mathrm{m}+\mathrm{s})\,[\exp(\mathrm{s})-1]$. Las estimaciones puntuales (*point estimates*) implican que los coeficientes relativos a la disponibilidad de peces, valoración estética y costo del viaje tienen las siguientes medianas, medias y desviaciones estándar:

Variable	Mediana	Media	Desv. Estándar
Disponibilidad de peces	0.0563	0.0944	0.1270
Valoración estética	0.4519	0.6482	0.6665
Costo del viaje	0.0906	0.1249	0.1185

El ratio para un pescador entre sus coeficiente de disponibilidad de peces y costo del viaje es una medida de la cantidad económica que el pescador está dispuesto a pagar para tener más peces en el río. Dado que el ratio entre dos términos distribuidos log-normal independientemente también se distribuye log-normal, podemos calcular los momentos estadísticos de la distribución de la predisposición a pagar. El logaritmo del ratio entre el coeficiente de disponibilidad de peces y de costo del viaje tiene una media estimada de -0.474 y una desviación estándar de 1.29. La relación en sí, por lo tanto, tiene una mediana de 0.62, media de 1.44 y desviación estándar de 2.96. Es decir, la predisposición media a pagar para que se incremente la disponibilidad de peces en 100 peces por cada 1000 pies de río se estima en 1.44$, y la variación en la predisposición de los pescadores a pagar por un disponibilidad de peces adicional es muy amplia. Del mismo modo, la predisposición media estimada a pagar por un sitio que tiene una valoración estética superior a 1 es de 9.87$, y de nuevo la variación es bastante grande.

Como ilustra este caso, el modelo logit mixto proporciona más información que un logit estándar, ya que el logit mixto estima hasta qué punto los pescadores difieren en sus preferencias por los atributos de los sitios. Las desviaciones estándar de los coeficientes son significativas, lo que indica que un logit mixto proporciona una mejor representación de la situación de elección que un logit estándar, que supone que los coeficientes son los mismos para todos los pescadores. El modelo logit mixto permite también contemplar varios viajes de cada pescador en la muestra y que las preferencias de cada pescador se apliquen a cada uno sus viajes.