

# 11

## Parámetros a nivel individual

### 11.1 Introducción

Los modelos logit mixto y probit permiten coeficientes aleatorios cuya distribución en la población se estima. Consideremos, por ejemplo, el modelo descrito el capítulo 6, relativo a la elección hecha por los pescadores entre distintos sitios de pesca disponibles. Los sitios se diferencian por el hecho de tener o no tener campings. A algunos pescadores les gusta tener campings en los sitios de pesca, ya que pueden utilizarlos para acampar y pasar la noche. A otros pescadores no les gustan las multitudes y el ruido que se asocian a los lugares para acampar, y prefieren la pesca en lugares más aislados. Para capturar estas diferencias en las preferencias, se especificó un modelo logit mixto que incluía coeficientes aleatorios tanto para la variable relativa a la presencia de camping como para otros atributos del sitio de pesca. Una vez especificado el modelo, se estimó la distribución de los coeficientes en la población. La figura 11.1 muestra la distribución estimada para el coeficiente de camping. Se especificó una distribución normal para dicho coeficiente. La media estimada fue de 0.116 y la desviación estándar de 1.655. Esta distribución proporciona información útil sobre la población. Por ejemplo, las estimaciones resultantes implican que al 47% de la población no le gusta tener lugares para acampar en sus sitios de pesca, mientras que al otro 53% por ciento sí le gusta.

La cuestión que se plantea es: ¿en qué parte de la distribución de preferencias está un pescador en particular? ¿Hay una manera de determinar si una persona dada tiende a preferir (o no), tener lugares para acampar en los sitios de pesca?

Las elecciones de una persona revelan algo acerca de sus gustos y preferencias, algo que el investigador puede, en principio, descubrir. Si el investigador observa que un pescador concreto escoge consistentemente sitios sin campamentos, aun cuando el costo de conducir hasta estos sitios es mayor, el investigador puede inferir razonablemente que a este pescador no le gusta acampar. Existe una manera precisa de llevar a cabo este tipo de inferencia, dada por Revelt y Train (2000).

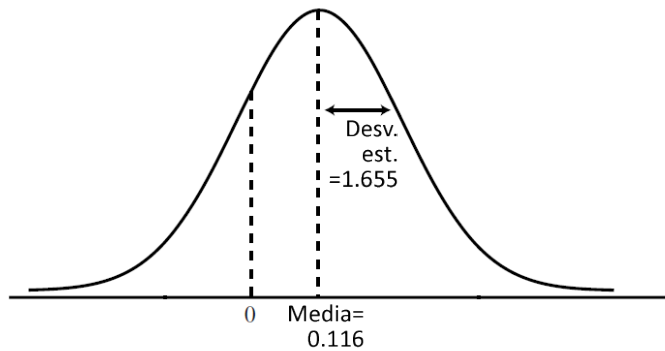


Figura 11.1. Distribución en el total de la población de pescadores del coeficiente relativo a la presencia de campings.

Explicamos el procedimiento en el contexto de un modelo logit mixto; sin embargo, el procedimiento puede usarse con cualquier modelo de comportamiento que incorpore coeficientes aleatorios, incluyendo probit. El concepto central es una distinción entre dos distribuciones: la distribución de las preferencias en la población y la distribución de las preferencias en la subpoblación de personas que toman unas decisiones concretas. Nos referiremos a los coeficientes aleatorios como al vector  $\beta$ . La distribución de  $\beta$  en el total de personas de la población se denota como  $g(\beta|\theta)$ , donde  $\theta$  son los parámetros de esta distribución, tales como la media y la varianza.

Una situación de elección consiste en varias alternativas descritas colectivamente por las variables  $x$ . Consideremos el siguiente experimento mental. Supongamos que todos los miembros de la población se enfrentan a la misma situación de elección descrita por las mismas variables  $x$ . Una parte de la población elegirá cada alternativa. Considere las personas que optan por la alternativa  $i$ . Las preferencias de estas personas no son todas iguales: existe una distribución de los coeficientes entre estas personas. Sea  $h(\beta|i, x, \theta)$  la distribución de  $\beta$  en la subpoblación de personas que, ante la situación de elección descrita por las variables  $x$ , elegiría la alternativa  $i$ . De esta forma,  $g(\beta|\theta)$  es la distribución de  $\beta$  en toda la población y  $h(\beta|i, x, \theta)$  la distribución de  $\beta$  en la subpoblación de personas que elegirían la alternativa  $i$  al enfrentarse a una situación de elección descrita por  $x$ .

Podemos generalizar la notación para permitir elecciones repetidas. Sea  $y$  la secuencia de elecciones en una serie de situaciones de elección que se describen colectivamente por las variables  $x$ . La distribución de los coeficientes en la subpoblación de personas que harían la secuencia de elecciones  $y$  al enfrentarse a las situaciones de elección descritas por  $x$  se denota como  $h(\beta|y, x, \theta)$ .

Observe que  $h(\cdot)$  está condicionada a  $y$ , mientras que  $g(\cdot)$  no lo está. A veces es útil referirse a  $h$  como la distribución condicionada y a  $g$  como la distribución no condicionada. Dos de estas distribuciones se muestran en la figura 11.2. Si no supiéramos nada sobre las elecciones pasadas de una persona, entonces lo mejor que podemos hacer para describir sus preferencias es decir que sus coeficientes se encuentran en algún lugar de  $g(\beta|\theta)$ . Sin embargo, si se ha observado que la persona hizo las elecciones  $y$  cuando se enfrentó a las situaciones de elección descritas por  $x$ , entonces sabemos que los coeficientes de esa persona están en la distribución  $h(\beta|y, x, \theta)$ . Dado que  $h$  es más estrecha que  $g$ , tenemos una mejor información sobre las preferencias de la persona al condicionar sobre sus elecciones pasadas.

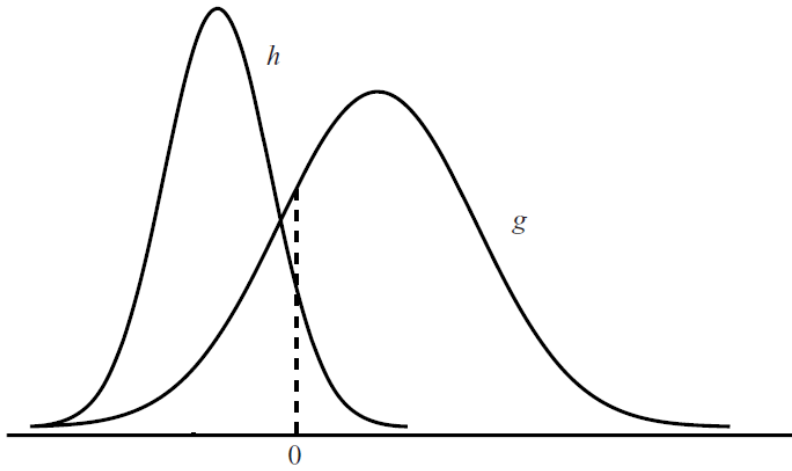


Figura 11.2. Distribución  $g$  (no condicionada) de la población y distribución  $h$  (condicionada) de la subpoblación de pescadores que eligen sitios de pesca sin campings.

Inferencias de este tipo hace mucho tiempo que se llevan a cabo en modelos de regresión lineal, donde tanto la variable dependiente y como la distribución de los coeficientes son continuas (Griffiths, 1972; Juez et al, 1988). Modelos de cambio de régimen (*regime-switching models*, series temporales que ocasionalmente muestran un cambio brusco en su comportamiento), sobre todo en macroeconomía, han utilizado un procedimiento análogo para evaluar la probabilidad de que una observación se encuentre dentro de un régimen determinado (Hamilton y Susmel, 1994 ; Hamilton, 1996). En estos modelos, la variable dependiente es continua y la distribución de los coeficientes es discreta (representando un conjunto de coeficientes para cada régimen). A diferencia de estos dos casos, nuestros modelos tienen variables dependientes discretas. Kamakura y Russell (1989) y DeSarbo et al. (1995) desarrollaron un enfoque del problema en el contexto de un modelo de elección discreta con una distribución discreta de los coeficientes (es decir, un modelo de clase latente). Utilizaron para ello procedimientos de máxima verosimilitud con el fin de estimar los coeficientes de cada segmento, y luego calcularon la probabilidad de que una observación estuviese dentro de cada segmento, basándose en las elecciones registradas para esa observación. El enfoque que se describe en este capítulo aplica a modelos de elección discreta con coeficientes distribuidos de forma continua o discreta, y utiliza máxima verosimilitud (u otros métodos clásicos) para la estimación. Los modelos de Kamakura y Russell (1989) y DeSarbo et al. (1995) son un caso especial de este método más general. También se han desarrollado procedimientos bayesianos para llevar a cabo esta inferencia en modelos de elección discreta (Rossi et al, 1996;. Allenby y Rossi, 1999). Los métodos bayesianos se describen en el capítulo 12.

## 11.2 Derivación de la distribución condicionada

La relación entre  $h$  y  $g$  se puede establecer con precisión. Considere una elección entre alternativas  $j = 1, \dots, J$  en diferentes situaciones de elección  $t = 1, \dots, T$ . La utilidad que la persona  $n$  obtiene al elegir la alternativa  $j$  en una situación  $t$  es

$$U_{njt} = \beta'_n x_{njt} + \varepsilon_{njt},$$

donde  $\varepsilon_{njt} \sim iid$  valor extremo y  $\beta_n \sim g(\beta|\theta)$  en la población. Podemos referirnos a las variables  $x_{njt}$  de forma colectiva, para todas las alternativas y todas las situaciones de elección, como  $x_n$ . Sea  $y_n = \langle y_{n1}, \dots, y_{nT} \rangle$  la secuencia de alternativas escogidas por la persona. Si conociésemos  $\beta_n$ , entonces la probabilidad de la secuencia de elecciones de la persona sería un producto de logits:

$$P(y_n|x_n, \beta) = \prod_{t=1}^T L_{nt}(y_{nt}|\beta),$$

donde

$$L_{nt}(y_{nt}|\beta) = \frac{e^{\beta' x_{ny_{nt}t}}}{\sum_j e^{\beta' x_{njt}}}.$$

Dado que no conocemos  $\beta_n$ , la probabilidad de la secuencia de elecciones de la persona es la integral de  $P(y_n|x_n, \beta)$  sobre la distribución de  $\beta$ :

$$(11.1) \quad P(y_n|x_n, \theta) = \int P(y_n|x_n, \beta) g(\beta|\theta) d\beta.$$

Esta es la probabilidad de elección de un modelo logit mixto que ya comentamos en el capítulo 6.

Ahora podemos obtener  $h(\beta|y_n, x_n, \theta)$ . Por el teorema de Bayes,

$$h(\beta|y_n, x_n, \theta) \times P(y_n|x_n, \theta) = P(y_n|x_n, \beta) \times g(\beta|\theta).$$

Esta ecuación simplemente establece que la densidad conjunta de  $\beta$  e  $y_n$  se puede expresar como la probabilidad de  $y_n$  por la probabilidad de  $\beta$  condicionada a  $y_n$  (que es el lado izquierdo de la expresión), o en la otra dirección del condicionamiento, como la probabilidad de  $\beta$  por la probabilidad de  $y_n$  condicionada a  $\beta$  (que es el lado derecho). Reordenando,

$$(11.2) \quad h(\beta|y_n, x_n, \theta) = \frac{P(y_n|x_n, \beta) \times g(\beta|\theta)}{P(y_n|x_n, \theta)}.$$

Sabemos todas las cantidades del lado derecho. A partir de estas cantidades, podemos calcular  $h$ .

La ecuación (11.2) también proporciona una forma de interpretar  $h$  intuitivamente. Observe que el denominador  $P(y_n|x_n, \theta)$  es la integral del numerador, tal y como se indica en la definición (11.1). Como tal, el denominador es una constante que hace que la integral de  $h$  sea 1, como se requiere para cualquier densidad. Dado que el denominador es una constante,  $h$  es proporcional al numerador,  $P(y_n|x_n, \beta) \times g(\beta|\theta)$ . Esta relación hace que la interpretación de  $h$  sea relativamente fácil. Dicho en palabras, la densidad de  $\beta$  en la subpoblación de personas que elegirían la secuencia  $y_n$  al enfrentarse a las situaciones de elección  $x_n$  es proporcional al producto de la densidad de  $\beta$  en la totalidad de la población por la probabilidad de que  $y_n$  sea elegida si los coeficientes de la persona fuesen  $\beta$ .

Usando (11.2), se pueden obtener varios estadísticos condicionados a  $y_n$ . La  $\beta$  media en la subpoblación de personas que elegirían  $y_n$  cuando se enfrentan a  $x_n$  es

$$\bar{\beta}_n = \int \beta \cdot h(\beta|y_n, x_n, \theta) d\beta.$$

Esta media generalmente difiere de la  $\beta$  media en toda la población. Sustituyendo  $h$  en la fórmula,

$$\bar{\beta}_n = \frac{\int \beta \cdot P(y_n|x_n, \beta) g(\beta|\theta) d\beta}{P(y_n|x_n, \theta)}$$

$$(11.3) \quad = \frac{\int \beta \cdot P(y_n | x_n, \beta) g(\beta | \theta) d\beta}{\int P(y_n | x_n, \beta) g(\beta | \theta) d\beta}.$$

Las integrales en esta ecuación no tienen una forma cerrada; sin embargo, se pueden simular fácilmente. Para ello, extraiga valores al azar de  $\beta$  a partir de la densidad poblacional  $g(\beta | \theta)$ . Calcule el promedio ponderado de estos valores, siendo el peso del valor  $\beta^r$  proporcional a  $P(y_n | x_n, \beta^r)$ . La media en la subpoblación simulada es

$$\check{\beta}_n = \sum_r \omega^r \beta^r,$$

donde los pesos son

$$(11.4) \quad \omega^r = \frac{P(y_n | x_n, \beta^r)}{\sum_r P(y_n | x_n, \beta^r)}.$$

Es posible calcular otros estadísticos. Supongamos que la persona se enfrenta a una nueva situación de elección descrita por variables  $x_{njT+1} \forall j$ . Si no tuviéramos información sobre las decisiones pasadas de la persona, podríamos asignar la siguiente probabilidad de que eligiera la alternativa i:

$$(11.5) \quad P(i | x_{nT+1}, \theta) = \int L_{nT+1}(i | \beta) g(\beta | \theta) d\beta,$$

donde

$$L_{nT+1}(i | \beta) = \frac{e^{\beta' x_{niT+1}}}{\sum_j e^{\beta' x_{njT+1}}}.$$

Esto simplemente es la probabilidad de un modelo logit mixto utilizando la distribución de  $\beta$  en la población. Sin embargo, si hemos observado las últimas elecciones de la persona, entonces la probabilidad puede condicionarse a estas elecciones. La probabilidad se convierte en

$$(11.6) \quad P(i | x_{nT+1}, y_n, x_n, \theta) = \int L_{nT+1}(i | \beta) h(\beta | y_n, x_n, \theta) d\beta.$$

Esta es también es la probabilidad del modelo logit mixto, pero utilizando la distribución condicionada  $h$  en lugar de la distribución no condicionada  $g$ . Cuando no conocemos las elecciones previas de la persona, combinamos la fórmula logit sobre la densidad de  $\beta$  en toda la población. Por el contrario, cuando sí conocemos las elecciones anteriores de la persona, podemos mejorar nuestra predicción mediante la combinación sobre la densidad de  $\beta$  en la subpoblación que habría hecho las mismas elecciones que esta persona.

Para calcular esta probabilidad, sustituimos  $h$  en la fórmula a partir de (11.2):

$$P(i | x_{nT+1}, y_n, x_n, \theta) = \frac{\int L_{nT+1}(i | \beta) P(y_n | x_n, \beta) g(\beta | \theta) d\beta}{\int P(y_n | x_n, \beta) g(\beta | \theta) d\beta}.$$

La probabilidad se simula extrayendo valores al azar  $\beta$  de la distribución  $g$  en la población, calculando la fórmula logit para cada valor, y tomando un promedio ponderado de los resultados:

$$\check{P}_{niT+1}(y_n, x_n, \theta) = \sum_r \omega^r L_{nT+1}(i|\beta^r),$$

donde los pesos están dados por (11.4).

### 11.3 Implicaciones de la estimación de $\theta$

Los parámetros poblacionales  $\theta$  se estiman usando cualquiera de los métodos que se describen en el capítulo 10. El enfoque más común es el de máxima verosimilitud simulada, usando el valor simulado de  $P(y_n|x_n, \theta)$  en la función log-verosimilitud. Obtenemos así una estimación de  $\theta$ , etiquetada  $\hat{\theta}$ . Sabemos que el estimador tiene varianza muestral. La covarianza asintótica del estimador también se estima, la cual denominamos  $\hat{W}$ . Por consiguiente, la distribución asintótica se estima como  $N(\hat{\theta}, \hat{W})$ .

El parámetro  $\theta$  describe la distribución de  $\beta$  en la población, dando, por ejemplo, la media y la varianza de  $\beta$  sobre todos los decisores. Para cualquier valor de  $\theta$ , la ecuación (11.2) nos da la distribución condicionada de  $\beta$  en la subpoblación de personas que harían las elecciones  $y_n$  al enfrentarse a situaciones de elección descritas por  $x_n$ . Esta relación es exacta en el sentido de que no hay ninguna varianza, muestral o de cualquier otro tipo, asociada a ella. De forma similar, cualquier estadístico basado en  $h$  es exacto dado un valor de  $\theta$ . Por ejemplo, la media de la distribución condicionada,  $\bar{\beta}_n$ , es exactamente la ecuación (11.3) para un valor dado de  $\theta$ .

Teniendo en cuenta esta correspondencia entre  $\theta$  y  $h$ , el hecho de que  $\theta$  sea un valor estimado se puede manejar de dos maneras diferentes. El primer enfoque consiste en usar la estimación puntual (*point estimate*) de  $\theta$  para calcular los estadísticos asociados con la distribución condicionada  $h$ . Bajo este enfoque, la media de la distribución condicionada,  $\bar{\beta}_n$ , se calcula insertando  $\hat{\theta}$  en (11.3). La probabilidad de una nueva situación de elección se calcula insertando  $\hat{\theta}$  en (11.6). Si el estimador de  $\theta$  es consistente, entonces este enfoque es consistente para estadísticos basados en  $\theta$ .

El segundo enfoque consiste en considerar la distribución muestral de  $\hat{\theta}$ . Cada posible valor de  $\theta$  implica un valor de  $h$ , y por lo tanto un valor de cualquier estadístico asociado con  $h$ , como  $\bar{\beta}_n$ . La varianza muestral en el estimador de  $\theta$  induce varianza muestral en los estadísticos que son calculados sobre la base de  $\theta$ . Esta varianza muestral se puede calcular mediante simulación, a través de la extracción de valores al azar de  $\theta$  a partir de su distribución muestral estimada y posterior cálculo del estadístico correspondiente para cada uno de estos valores.

Por ejemplo, para representar la distribución muestral de  $\hat{\theta}$  en el cálculo de  $\bar{\beta}_n$ , se siguen los siguientes pasos:

1. Extraiga un valor al azar de  $N(\hat{\theta}, \hat{W})$ , que es la distribución muestral estimada de  $\hat{\theta}$ . Este paso se lleva a cabo de la siguiente manera. Extraiga  $K$  valores al azar de una densidad normal estándar, y etiquete el vector de estos valores como  $\eta^r$ , donde  $K$  es la longitud de  $\theta$ . A continuación, cree  $\theta^r = \hat{\theta} + L\eta^r$ , donde  $L$  es el factor Choleski de  $\hat{W}$ .
2. Calcule  $\bar{\beta}_n^r$  basándose en este  $\theta^r$ . Dado que la fórmula para  $\bar{\beta}_n$  implica integración, la simulamos usando la fórmula (11.3).
3. Repita los pasos 1 y 2 múltiples veces, etiquetando el número de veces como  $R$ .

Los valores resultantes son extracciones de la distribución muestral de  $\bar{\beta}_n$  inducida por la distribución muestral de  $\hat{\theta}$ . La media de  $\bar{\beta}_n^r$  sobre los  $R$  sorteos de  $\theta^r$  es la media de la distribución muestral de  $\bar{\beta}_n$ . La desviación estándar de los valores extraídos al azar da el error estándar asintótico de  $\bar{\beta}_n$  que es inducido por la varianza muestral de  $\hat{\theta}$ .

Tenga en cuenta que este proceso implica simulación dentro de una simulación. Para cada valor al azar de  $\theta^r$ , el estadístico  $\bar{\beta}_n^r$  es simulado con múltiples sorteos de  $\beta$  extraídos de la densidad poblacional  $g(\beta|\theta^r)$ .

Supongamos que usamos cualquiera de estos enfoques para estimar  $\bar{\beta}_n$ . A continuación surge la siguiente pregunta: ¿puede considerarse la estimación de  $\bar{\beta}_n$  una estimación de  $\beta_n$ ? Es decir: ¿es la media estimada de la distribución condicionada  $h(\beta|y_n, x_n, \theta)$ , que está condicionada a decisiones pasadas de la persona  $n$ , una estimación de los coeficientes de la persona  $n$ ?

Hay dos respuestas posibles, dependiendo de cómo ve el investigador el proceso de generación de datos. Si el número de situaciones de elección que el investigador puede observar para cada decisor es fijo, entonces la estimación de  $\bar{\beta}_n$  no es una estimación consistente de  $\beta_n$ . Cuando  $T$  es fijo, la consistencia exige que la estimación converja al verdadero valor cuando el tamaño de la muestra aumenta sin límite. Si el tamaño de la muestra aumenta, pero las situaciones de elección que enfrenta la persona  $n$  son fijas, entonces la distribución condicionada y su media no cambian. En la medida en que los coeficientes de la persona  $n$  no coinciden con la media de la distribución condicionada (una circunstancia esencialmente imposible), la media de la distribución condicionada nunca será igual a los coeficientes de la persona, sin importar cómo de grande sea la muestra. Elevar el tamaño de la muestra mejora la estimación de  $\theta$  y por lo tanto proporciona una mejor estimación de la media de la distribución condicionada, ya que esta media sólo depende de  $\theta$ . Sin embargo, aumentar el tamaño de la muestra no hace que la media condicionada sea igual a los coeficientes de la persona.

Cuando se fija el número de situaciones de elección, la media condicionada tiene la misma interpretación que la media de la población, pero para un grupo de gente diferente y menos diversa. Cuando predecimos el comportamiento futuro de la persona, es de esperar que obtengamos mejores predicciones usando la distribución condicionada, como en (11.6), que la distribución en la población. En el estudio de un caso que se presenta en la siguiente sección, se muestra que la mejora puede ser sustancial.

Si se puede considerar que el número de situaciones de elección a las que una persona se enfrenta crece, entonces la estimación de  $\bar{\beta}_n$  puede ser considerada como una estimación de  $\beta_n$ . Sea  $T$  el número de situaciones de elección que la persona  $n$  afronta. Si observamos más elecciones de la persona (es decir,  $T$  se eleva), entonces estamos en mejores condiciones para identificar los coeficientes de la persona. La figura 11.3 muestra la distribución condicionada  $h(\beta|y_n, x_n, \theta)$  para tres valores diferentes de  $T$ . La distribución condicionada tiende a moverse hacia la  $\beta_n$  de la propia persona a medida que aumenta  $T$ , y tiende a ser más concentrada. A medida que  $T$  aumenta sin límite, la distribución condicionada colapsa en  $\beta_n$ . La media de la distribución condicionada converge al verdadero valor de  $\beta_n$  a medida que el número de situaciones de elección aumenta sin límite. Por tanto, la estimación de  $\bar{\beta}_n$  es consistente para  $\beta_n$ .

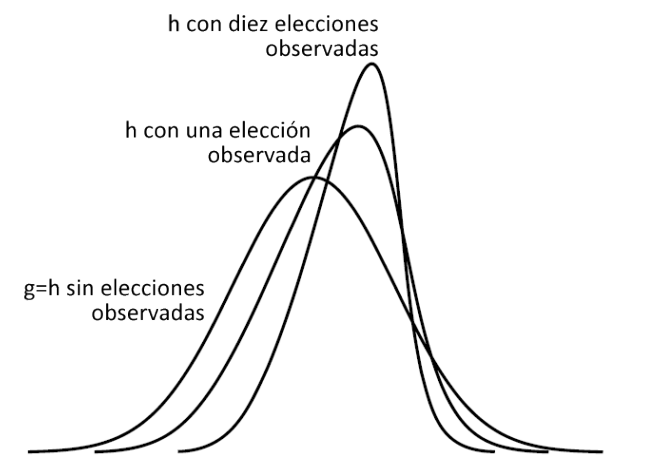


Figura 11.3. Distribución condicionada con  $T=0, 1$  y  $10$ .

En el Capítulo 12 se describe el teorema de Bernstein-von Mises. Este teorema indica que, en condiciones bastante laxas, la media de una distribución posterior de un parámetro es asintóticamente equivalente al máximo de la función de verosimilitud. La distribución condicionada  $h$  es una distribución posterior: por (11.2)  $h$  es proporcional a una densidad  $g$ , que puede ser interpretada como una distribución a priori de  $\beta_n$ , multiplicada por la verosimilitud de las  $T$  elecciones hechas por la persona  $n$  dadas las  $\beta_n$ , que es  $P(y_n|x_n, \beta_n)$ . Por el teorema de Bernstein-von Mises, la media de  $h$  es por lo tanto un estimador de  $\beta_n$  que es asintóticamente equivalente al estimador de máxima verosimilitud de  $\beta_n$ , donde el comportamiento asintótico se define como un aumento de  $T$ . Estos conceptos se describen más detalladamente en el capítulo 12; los mencionamos ahora simplemente para proporcionar otra interpretación de la media de la distribución condicionada.

#### 11.4 Ilustración de Monte Carlo

Para ilustrar los conceptos expuestos, generé un conjunto de datos hipotéticos donde los verdaderos parámetros de la población  $\theta$  son conocidos, así como el verdadero  $\beta_n$  de cada decisor. Estos datos nos permiten comparar la media de la distribución condicionada de las elecciones de cada decisor,  $\bar{\beta}_n$ , con la  $\beta_n$  de cada decisor. También nos permiten investigar el impacto de un aumento del número de situaciones de elección en la distribución condicionada. Para este experimento, construí conjuntos de datos que constaban de 300 "clientes", cada uno enfrentando  $T = 1, 10, 20$  y 50 situaciones de elección. Cada conjunto de datos consta de tres alternativas y cuatro variables. Los coeficientes para las dos primeras variables se mantienen fijos para toda la población, con valor 1.0, y los coeficientes de las dos últimas variables se distribuyen normalmente con media y varianza 1.0. La utilidad se especifica de forma que incluye estas variables más un término final iid que se distribuye valor extremo, de manera que el modelo es logit mixto. La variable dependiente para cada cliente se ha creado extrayendo un valor al azar de las densidades de los términos aleatorios, calculando la utilidad de cada alternativa con este valor, y determinando cuál de las alternativas tenía la utilidad más alta. Para minimizar el efecto del ruido de simulación en la creación de los datos, construí 50 conjuntos de datos para cada nivel de  $T$ . Los resultados que se reportan son el promedio entre estos 50 conjuntos de datos.

Calculamos la media de la distribución condicionada para cada cliente,  $\bar{\beta}_n$ . También calculamos la desviación estándar de  $\bar{\beta}_n$  entre los 300 clientes, así como la desviación media absoluta de  $\bar{\beta}_n$  respecto a las  $\beta_n$  de los clientes (es decir, la media sobre  $n$  de  $|\bar{\beta}_n - \beta_n|$ ). La tabla 11.1 presenta estos estadísticos. Consideremos en primer lugar la desviación estándar. Si no hubiera situaciones de elección observadas sobre las que condicionar ( $T = 0$ ), entonces la distribución condicionada para cada cliente sería igual a la distribución no condicionada (población). Cada cliente tendría la misma  $\bar{\beta}_n$ , igual a la media de  $\beta_n$  en la población. En este caso, la desviación estándar de  $\bar{\beta}_n$  sería cero, ya que todos los clientes tendrían la misma  $\bar{\beta}_n$ . En el otro extremo, si se observó un número ilimitadamente grande de situaciones de elección ( $T \rightarrow \infty$ ), entonces la distribución condicionada para cada cliente colapsaría en su propia  $\beta_n$ . En este caso, la desviación estándar de  $\bar{\beta}_n$  sería igual a la desviación estándar de la distribución de  $\beta_n$  en la población, que es 1.0 en este experimento. Para  $T$  entre 0 y  $\infty$ , la desviación estándar de  $\bar{\beta}_n$  está entre 0 y la desviación estándar de  $\beta_n$  en la población.

Tabla 11.1. Ilustración de Monte Carlo

	Coeficiente 1º	Coeficiente 2º
1 situación de elección:		
Desv. Estándar de $\bar{\beta}_n$	0.413	0.416



Diferencia absoluta entre $\bar{\beta}_n$ y $\beta_n$	0.726	0.718
10 situación de elección:		
Desv. Estándar de $\bar{\beta}_n$	0.826	0.826
Diferencia absoluta entre $\bar{\beta}_n$ y $\beta_n$	0.422	0.448
20 situación de elección:		
Desv. Estándar de $\bar{\beta}_n$	0.894	0.886
Diferencia absoluta entre $\bar{\beta}_n$ y $\beta_n$	0.354	0.350
50 situación de elección:		
Desv. Estándar de $\bar{\beta}_n$	0.951	0.953
Diferencia absoluta entre $\bar{\beta}_n$ y $\beta_n$	0.243	0.243

En la tabla 11.1 vemos que condicionar sobre pocas situaciones de elección captura una gran parte de la variación de  $\beta$  entre clientes. Con sólo una situación de elección, la desviación estándar de  $\bar{\beta}_n$  está en torno a 0.4. Puesto que la desviación estándar de  $\beta_n$  en la población es de 1.0 en este experimento, esto significa que condicionar respecto a una situación de elección captura más del 40 por ciento de la variación en  $\beta_n$ . Con 10 situaciones de elección, capturamos más del 80 de la variación. El rendimiento de observar más situaciones de elección decrece fuertemente. Doblar  $T$  de  $T = 10$  a  $T = 20$  sólo aumenta la proporción de variación capturada desde aproximadamente 0.83 a aproximadamente 0.89. El aumento a  $T = 50$  aumenta la proporción a 0.95.

Consideremos ahora la diferencia absoluta entre la media de la distribución condicionada a las elecciones del cliente,  $\bar{\beta}_n$ , y la  $\beta_n$  real del cliente. Sin condicionar ( $T = 0$ ), la diferencia absoluta media sería de 0.8, que es la diferencia absoluta esperada para variables que siguen una distribución normal estándar como las que tenemos en nuestro experimento. Con un condicionamiento perfecto ( $T \rightarrow \infty$ ),  $\bar{\beta}_n = \beta_n$  para cada cliente, por lo que la diferencia absoluta sería de 0. Condicionando a una única situación de elección, la desviación absoluta media cae de 0.8 (sin condicionar) a aproximadamente 0.72, es decir, mejora un 10%. La desviación absoluta va decreciendo a medida que el número de situaciones de elección se eleva.

Observe que la caída de la desviación absoluta es menor que el aumento en la desviación estándar. Por ejemplo, con una única situación de elección, la desviación absoluta se mueve un 10% desde su valor en ausencia de condicionamiento hacia el valor con un conocimiento perfecto (de 0.80 con  $T = 0$  a 0.72 con  $T = 1$ , que es un 10% del recorrido entre no condicionar,  $T = 0$ , y tener un conocimiento perfecto,  $T \rightarrow \infty$ ). Sin embargo, la desviación estándar se mueve alrededor de un 40% desde su valor en ausencia de condicionamiento y su valor con conocimiento perfecto (0.4 con  $T = 1$  es el 40% de la distancia desde 0 con  $T = 0$  a 1 con  $T \rightarrow \infty$ ). Esta diferencia se debe al hecho de que la desviación estándar incorpora tanto un movimiento de  $\bar{\beta}_n$  en dirección contraria a  $\beta_n$  (alejamiento) como un movimiento hacia  $\beta_n$  (acercamiento). Es importante tener en cuenta este hecho cuando se evalúa la desviación estándar de  $\bar{\beta}_n$  en aplicaciones empíricas, en las que la diferencia absoluta no se puede calcular al desconocerse la  $\beta_n$ . Es decir, la desviación estándar de  $\bar{\beta}_n$  expresada como un porcentaje de la desviación estándar estimada en la población es una sobreestimación de la cantidad de información contenida en las  $\bar{\beta}_n$ 's. Con diez situaciones de elección, la desviación estándar promedio en  $\bar{\beta}_n$  es más del 80% del valor que tendría con un conocimiento perfecto, y sin embargo la desviación absoluta es menos de la mitad de la que sería sin condicionar.

### 11.5 Distribución condicionada promedio

Para un modelo correctamente especificado en los verdaderos parámetros de la población, la distribución condicionada de las preferencias, agregada para todos los clientes, es igual a la distribución poblacional de las preferencias. Dada una serie de situaciones de elección descritas por  $x_n$ , hay un

conjunto de posibles secuencias de elección. Etiquetemos estas posibles secuencias como  $y_s$  para  $s = 1, \dots, S$ . Llamemos a la verdadera frecuencia de  $y_s$  como  $m(y_s|x_n, \theta^*)$ , explicitando su dependencia de los verdaderos parámetros  $\theta^*$ . Si se ha especificado correctamente el modelo y se ha estimado consistentemente, entonces  $P(y_s|x_n, \hat{\theta})$  aproxima  $m(y_s|x_n, \theta^*)$  asintóticamente. Condicionando respecto a las variables explicativas, el valor esperado de  $h(\beta|y_s, x_n, \hat{\theta})$  es

$$E_y h(\beta|y, x_n, \hat{\theta}) = \sum_s \frac{P(y_s|x_n, \beta)g(\beta|x_n, \hat{\theta})}{P(y_s|x_n, \hat{\theta})} m(y_s|x_n, \theta^*)$$

$$\rightarrow \sum_s P(y_s|x_n, \beta)g(\beta|x_n, \hat{\theta}) = g(\beta|x_n, \hat{\theta}).$$

Esta relación proporciona una herramienta de diagnóstico del modelo (Allenby y Rossi, 1999). Si la media de las distribuciones condicionadas de las preferencias en los clientes muestreados es similar a la distribución estimada en la población, el modelo está correctamente especificado y calculado con precisión. Si no son similares, la diferencia podría ser debida a (1) un error de especificación, (2) un número insuficiente de valores aleatorios empleados en la simulación, (3) un tamaño de muestra inadecuado y/o (4) la rutina de máxima verosimilitud ha convergido en un máximo local en lugar de un máximo global.

## 11.6 Caso de estudio: elección de proveedor de energía

### 11.6.1 Distribución en la población

Se obtuvieron datos de preferencia declarada relativa a la elección del proveedor de electricidad de clientes residenciales. A los clientes encuestados se les presentaron 8-12 situaciones de elección hipotéticas llamadas experimentos. En cada experimento, al cliente se le presentaron cuatro proveedores alternativos con diferentes precios y otras características. En cuanto a precios, los proveedores ofrecían tres tipos de tarifas: (a) precio fijo en centavos de dólar por kilovatio-hora, c/kWh, (b) precios “time-of-day” (TOD), variables en función de tramos horarios o (c) precios estacionales con precios diferentes para cada estación del año. El resto de características eran la duración del contrato (tiempo durante el cual el proveedor debe proporcionar el servicio al precio contratado y el cliente debe permanecer con el proveedor si no desea pagar una sanción por haberlo abandonado antes), así como el hecho de si el proveedor es local, una empresa conocida aunque no sea local o una empresa poco conocida. Los datos fueron recogidos por *Research Triangle Institute* (1997) para el *Electric Power Research Institute* y fueron utilizados por Goett (1998) para estimar modelos logits mixtos. Utilizaremos una especificación similar a la de Goett, pero hemos eliminado o combinado variables que resultaron ser insignificantes.

Se estimaron dos modelos logit mixto con estos datos, basados en diferentes especificaciones para la distribución de los coeficientes aleatorios. Todas las elecciones excepto la última situación de elección de cada cliente se utilizaron para estimar los parámetros de la distribución en la población, y la última situación de elección se reservó para comparar la capacidad predictiva de los diferentes modelos y métodos.

La tabla 11.2 muestra los parámetros estimados en la población. El coeficiente de precio en los dos modelos se fija en toda la población de tal manera que la distribución de la predisposición a pagar por cada atributo distinto del precio (que es el ratio entre el coeficiente del atributo en cuestión y el coeficiente de precio) tiene la misma distribución que el coeficiente del propio atributo. Para el modelo 1, todos los coeficientes distintos del precio se especifican mediante una distribución normal en la

población. Se estima la media  $m$  y la desviación estándar  $s$  de cada coeficiente. Para el modelo 2, los tres primeros coeficientes distintos del precio se especifican para que sean normales, y el cuarto y quinto para que sean log-normales. Las variables cuarta y quinta se refieren a las tarifas TOD y estacionales, y sus coeficientes lógicamente deberían ser negativos para todos los clientes. La distribución log-normal (con los signos de las variables invertidos) proporcionan este comportamiento. El logaritmo de estos coeficientes se distribuye normalmente con media  $m$  y desviación estándar  $s$ , que son los parámetros que realmente se estiman. Los coeficientes resultantes tienen media  $\exp(m + (s^2/2))$  y desviación estándar igual a la media por  $\sqrt{\exp(s^2) - 1}$ .

Tabla 11.2. Modelo logit mixto de elección de proveedor de energía

	Modelo 1	Modelo 2
Precio, kWh	-0.8574 (0.0488)	-0.8827 (0.0497)
Duración del contrato, años		
m	-0.1833 (0.0289)	-0.2125 (0.0261)
s	0.3786 (0.0291)	0.3865 (0.0278)
Compañía local		
m	2.0977 (0.1370)	2.2297 (0.1266)
s	1.5585 (0.1264)	1.7514 (0.1371)
Compañía conocida		
m	1.5247 (0.1018)	1.5906 (0.0999)
s	0.9520 (0.0998)	0.9621 (0.0977)
Tarifas TOD(a)		
m	-8.2857 (0.4577)	2.1328 (0.0543)
s	2.5742 (0.1676)	0.4113 (0.0397)
Tarifas estacionales(b)		
m	-8.5303 (0.4468)	2.1577 (0.0509)
s	2.1259 (0.1604)	0.2812 (0.0217)
Log-verosimilitud en convergencia	-3646.51	-3618.92

Errores estándar entre paréntesis

(a) Tarifas TOD: 11c/kWh, 8 a.m.–8 p.m., 5c/kWh, 8 p.m.–8 a.m.

(b) Tarifas estacionales: 10c/kWh en verano, 8c/kWh en invierno, 6c/kWh en primavera y otoño.

Las estimaciones proporcionan los siguientes resultados cualitativos:

- El cliente medio está dispuesto a pagar alrededor de 1/5 a 1/4 c/kWh más, dependiendo del modelo, con el fin de tener un contrato que sea un año más corto. Dicho a la inversa, un

proveedor que requiera a los clientes firmar un contrato de cuatro a cinco años debe descontar de su precio 1 c/kWh para atraer al cliente medio.

- Existe una considerable variación en las actitudes de los clientes hacia la duración del contrato, con una parte relevante de clientes que prefieren un contrato más largo que corto. Un contrato a largo plazo constituye un seguro para el cliente frente a los aumentos de precios, ya que el proveedor está obligado a respetar el precio establecido durante la duración del contrato. Este tipo de contratos, sin embargo, impiden al cliente beneficiarse de precios más bajos que puedan surgir durante la vigencia del mismo. Al parecer, muchos clientes valoran más la seguridad frente a un incremento de los precios que el riesgo de perder una oportunidad de disfrutar precios más bajos. El grado de heterogeneidad de los clientes implica que el mercado puede soportar contratos de diferentes duraciones, con proveedores obteniendo beneficios mediante la oferta de diferentes tipos de contratos dirigidos a diferentes segmentos de población.
- El cliente medio está dispuesto a pagar la friolera de 2.5 c/kWh más por su proveedor local que por un proveedor desconocido. Sólo una pequeña parte de los clientes prefieren un proveedor desconocido a su compañía eléctrica local. Este hallazgo tiene implicaciones importantes para la competencia. Implica que la entrada en el mercado residencial de proveedores previamente desconocidos será muy difícil, sobre todo porque los descuentos en los precios que los competidores pueden ofrecer en la mayoría de los mercados son bastante reducidos. La experiencia en California, donde sólo el 1 por ciento de los clientes residenciales han abandonado a su proveedor local después de varios años de libre competencia, es consistente con este hallazgo.
- El cliente medio está dispuesto a pagar 1.8 c/kWh más por un proveedor conocido que por uno desconocido. Los valores estimados de  $s$  implican que una parte considerable de los clientes estarían dispuestos a pagar más por un proveedor conocido que por su proveedor local, presumiblemente a causa de una mala experiencia o una actitud negativa hacia la compañía eléctrica local. Estos resultados implican que las empresas que son conocidas por los clientes, tales como los operadores de larga distancia, los operadores de telecomunicaciones locales, las empresas locales de cable e incluso los minoristas como Sears y Home Depot, pueden ser más exitosas atrayendo clientes para el suministro de electricidad que empresas desconocidas antes de su entrada en el mercado como proveedores de energía.
- El cliente promedio valora las tarifas TOD de una forma bastante consistente con los patrones de uso de este tipo de tarifa. En el modelo 1, el coeficiente medio de la variable indicadora de tarifa TOD implica que el cliente promedio considera que estas tarifas son equivalentes a un precio fijo de 9.7 c/kWh. En el modelo 2, la media estimada y la desviación estándar del logaritmo del coeficiente implican que la mediana de la predisposición a pagar es de 8.4 c/kWh y la media de 10.4 c/kWh, que abarca la media del modelo 1. Aquí 9.5 c/kWh es el precio medio que un cliente pagaría en un sistema tarifario TOD si el 75 por ciento de su consumo se produjese durante el día (entre las 8 de la mañana y las 8 de la tarde) y el otro 25 por ciento se produjese durante la noche. Estas proporciones, aunque tal vez un poco altas para el día, son razonables. Los valores estimados de  $s$  son altamente significativos, reflejando heterogeneidad en los patrones de uso y tal vez capacidad de los clientes para desplazar el horario de su consumo en respuesta a las tarifas TOD. Estos valores son mayores de lo que razonablemente cabe esperar, lo que implica que una parte no despreciable de los clientes tratan las tarifas TOD como equivalentes a un precio fijo más alto que el precio de la tarifa TOD más alto, o menor que el precio de la tarifa TOD más bajo.

- El cliente medio parece evitar tarifas estacionales por razones que van más allá de los propios precios. El cliente medio trata las tarifas estacionales como equivalentes a un precio fijo de 10 c/kWh, que es el precio más alto dentro de la tarifa estacional. Una posible explicación de este resultado se relaciona con la variación estacional de las facturas de los clientes. En muchas regiones, el consumo de electricidad es más alto en verano, cuando se usa el aire acondicionado, y las facturas son, por tanto, mayores en verano que en otras épocas del año, incluso en sistemas tarifarios fijos. La variación en las facturas entre diferentes meses, sin variación proporcional en los ingresos, hace que sea más difícil para los clientes pagar sus facturas en verano. De hecho, la falta de pago de la mayoría de los servicios de energía es más frecuente en verano. Las tarifas estacionales, que aplican un precio más alto en verano, incrementan la variación del importe de las facturas entre diferentes estaciones. Los clientes estarían evitando racionalmente un sistema tarifario que exacerba una dificultad ya existente. Si esta interpretación es correcta, entonces las tarifas estacionales combinadas con una facturación suavizada (en la cual el proveedor carga una parte de la factura de verano en la factura de invierno) podrían proporcionar una solución atractiva para los clientes y proveedores por igual.

El modelo 2 alcanza un valor de log-verosimilitud mayor que el modelo 1, presumiblemente debido a que la distribución log-normal asegura coeficientes negativos para las tarifas TOD y las tarifas estacionales.

#### 11.6.2 Distribuciones condicionadas

A continuación usaremos los modelos estimados para calcular las distribuciones condicionadas de los clientes y las medias de estas distribuciones. Calcularemos  $\bar{\beta}_n$  para cada cliente de dos maneras. En primer lugar, calculamos  $\bar{\beta}_n$  utilizando la ecuación (11.3) con las estimaciones puntuales de los parámetros de la población. En segundo lugar, utilizamos el procedimiento descrito en la sección 11.3 para integrar sobre la distribución muestral de los parámetros estimados de la población.

Las medias y las desviaciones estándar de  $\bar{\beta}_n$  entre clientes de la muestra, calculadas mediante estos dos métodos, se muestran en las tablas 11.3 y 11.4, respectivamente. El coeficiente de precio no se muestra en la tabla 11.3, pues es fijo para toda la población. La tabla 11.4 incorpora la distribución muestral de los parámetros en la población, que incluye la varianza en el coeficiente de precio.

Tabla 11.3.  $\bar{\beta}_n$  medio usando estimaciones puntuales de  $\hat{\theta}$ .

	Modelo 1	Modelo 2
Duración del contrato		
Media	-0.2028	-0.2149
Desviación estándar	0.3175	0.3262
Proveedor local		
Media	2.1205	2.2146
Desviación estándar	1.2472	1.3836
Compañía conocida		
Media	1.5360	1.5997
Desviación estándar	0.6676	0.6818
Tarifa TOD		
Media	-8.3194	-9.2584
Desviación estándar	2.2725	3.1051
Tarifa estacional		
Media	-8.6394	-9.1344
Desviación estándar	1.7072	2.0560

Tabla 11.4  $\bar{\beta}_n$ , medio usando la distribución muestral de  $\hat{\theta}$ 

	Modelo 1	Modelo 2
Precio		
Media	-0.8753	-0.8836
Desviación estándar	0.5461	0.0922
Duración del contrato		
Media	-0.2004	-0.2111
Desviación estándar	0.3655	0.3720
Proveedor local		
Media	2.1121	2.1921
Desviación estándar	1.5312	1.6815
Compañía conocida		
Media	1.5413	1.5832
Desviación estándar	0.9364	0.9527
Tarifa TOD		
Media	-9.1615	-9.0216
Desviación estándar	2.4309	3.8785
Tarifa estacional		
Media	-9.4528	-8.9408
Desviación estándar	1.9222	2.5615

Observe en primer lugar los resultados de la tabla 11.3. La media de  $\bar{\beta}_n$  está muy cerca de la media estimada de la población que aparece en la tabla 11.2. Esta similitud es la esperada para un modelo correctamente especificado y consistentemente estimado. La desviación estándar de  $\bar{\beta}_n$  sería cero si no hubiera condicionamiento y sería igual a la desviación estándar de la población, si el coeficiente de cada cliente se conociese con exactitud. Las desviaciones estándar de la tabla 11.3 están considerablemente por encima de cero y están bastante cerca de las desviaciones estándar estimadas para la población, mostradas en la tabla 11.2. Por ejemplo, en el modelo 1, la media condicionada del coeficiente de duración de contrato tiene una desviación estándar entre clientes de 0.318, y la estimación puntual de la desviación estándar en la población es 0.379. Por lo tanto, la variación en  $\bar{\beta}_n$  captura más del 70 por ciento de la variación total estimada en este coeficiente. Se obtienen resultados similares para otros coeficientes. Este resultado implica que la media de la distribución condicionada de un cliente captura una proporción bastante grande de la variación en los coeficientes entre clientes y tiene el potencial de ser útil para diferenciar clientes.

Tal y como vimos en la sección 11.5, es posible hacer un diagnóstico relativo a la bondad de la especificación y la estimación del modelo, mediante la comparación de la media muestral de las distribuciones condicionadas con la distribución poblacional estimada. Las medias de la tabla 11.3 representan las medias del promedio muestral de las distribuciones condicionadas. La desviación estándar de la distribución condicionada promediada en la muestra depende de la desviación estándar de  $\bar{\beta}_n$ , que se muestra en la tabla 11.3, más la desviación estándar de  $\beta_n - \bar{\beta}_n$ . Cuando se añade este último componente, la desviación estándar de cada coeficiente coincide muy aproximadamente con la desviación estándar estimada en la población. Esta equivalencia sugiere que no hay un error significativo de especificación y que los parámetros estimados en la población son bastante exactos. Esta sugerencia se ve matizada, sin embargo, por los resultados mostrados en la tabla 11.4.

La tabla 11.4 muestra la media muestral y la desviación estándar de la media de la distribución muestral de  $\bar{\beta}_n$ , que se induce por la distribución muestral de  $\hat{\theta}$ . Las medias de la tabla 11.4 son las medias del promedio en la muestra de  $h(\beta|y_n, x_n, \hat{\theta})$  integradas respecto a la distribución muestral de  $\hat{\theta}$ . Para el

modelo 1, se produce una discrepancia que podría indicar un posible error de especificación. En particular, las medias de los coeficientes relativos a las tarifas TOD y las tarifas estacionales de la tabla 11.4 exceden su media poblacional estimada, mostrada en la tabla 11.2. Curiosamente, las medias para estos coeficientes en la tabla 11.4 para el modelo 1 están más cerca de las medias análogas del modelo 2 que de las medias poblacionales estimadas para el modelo 1, mostradas en la tabla 11.2. El modelo 2 usa la distribución log-normal, cuya forma es más razonable para estos coeficientes, y obtiene un ajuste considerablemente mejor que el del modelo 1. El condicionamiento en el modelo 1 parece que desplaza los coeficientes hacia los valores obtenidos por el modelo 2 - mejor especificado - y los aleja de sus propias distribuciones poblacionales, deficientemente especificadas. Éste es un ejemplo de cómo comparar la distribución estimada de la población con el promedio en la muestra de la distribución condicionada puede revelar información acerca de la especificación y la estimación.

Las desviaciones estándar mostradas en la tabla 11.4 son más grandes que las de la tabla 11.3. Esta diferencia se debe al hecho de que la varianza muestral de los parámetros estimados en la población está incluida en los cálculos hechos para la tabla 11.4, pero no para la tabla 11.3. Las desviaciones estándar más grandes no significan que la porción de la varianza total de  $\beta_n$  que se captura por la variación en  $\bar{\beta}_n$  sea mayor cuando se tiene en cuenta la distribución en la muestra respecto a cuando no se tiene en cuenta.

Es posible obtener información útil para acciones de marketing mediante el examen de la  $\bar{\beta}_n$  de cada cliente. El valor de esta información para la comercialización segmentada de productos/servicios se ha destacado por Rossi et al. (1996). La tabla 11.5 muestra la  $\bar{\beta}_n$  calculada para los tres primeros clientes del conjunto de datos, junto con la media de  $\beta_n$  en la población.

*Tabla 11.5. Medias condicionadas para tres clientes*

	Población	Cliente 1	Cliente 2	Cliente 3
Duración del contrato	-0.213	0.198	-0.208	-0.401
Proveedor local	2.23	2.91	2.17	0.677
Compañía conocida	1.59	1.79	2.15	1.24
Tarifa TOD	-9.19	-5.59	-8.92	-12.8
Tarifa estacional	-9.02	-5.86	-11.1	-10.9

El primer cliente prefiere un contrato a largo plazo, a diferencia de la gran mayoría de los clientes, a los cuales no les gustan los contratos a largo plazo. Él está dispuesto a pagar un precio superior por la energía si este precio está garantizado a través de un contrato de larga duración. Este cliente evalúa las tarifas TOD y las tarifas estacionales de forma generosa, como si todo su consumo se produjese en el período de menor precio (tenga en cuenta que el precio más bajo con tarifa TOD es de 5 c/kWh y el precio más bajo con tarifa estacional es de 6 c/kWh). Es decir, el primer cliente probablemente está dispuesto a pagar por tener asignada una tarifa TOD o una tarifa estacional, más de lo que las tarifas realmente valen en términos de reducción de la factura. Por último, este cliente está dispuesto a pagar más que el promedio de los clientes por permanecer con la compañía eléctrica local. Desde una perspectiva de marketing, la compañía eléctrica local fácilmente puede retener y obtener beneficios adicionales de este cliente si le ofrece un contrato a largo plazo con tarifas TOD o tarifas estacionales.

Al tercer cliente no le gustan las tarifas estacionales y las tarifas TOD, evaluando ambas como si todo su consumo se produjese en los periodos con precios más caros. A este cliente le desagradan los contratos de larga duración mucho más que a la media de los clientes y, sin embargo, a diferencia de la mayoría de los clientes, prefiere recibir el servicio de una empresa conocida que no sea su compañía local. Este

cliente es un objetivo prioritario para una compañía bien conocida si esta compañía le ofrece un precio fijo sin requerirle un compromiso de permanencia.

El segundo cliente no es una oportunidad de marketing tan clara. Una empresa bien conocida está aproximadamente en igualdad de condiciones con la empresa local de energía para atraer a este cliente. Esto en sí mismo puede hacer que el cliente sea un objetivo para proveedores reconocidos, ya que está menos ligado a la empresa local de energía que la mayoría de los clientes. Sin embargo, más allá de esta información, hay poco más destacable en las preferencias de este cliente, salvo los precios bajos (que son valorados por todos los clientes). Su evaluación de las tarifas TOD y las tarifas estacionales es suficientemente negativa como para hacer poco probable que un proveedor pueda atraerlo y obtener un beneficio de él ofreciéndole estas tarifas. Asimismo, el cliente está dispuesto a pagar para evitar un contrato a largo plazo, por lo que un proveedor podría atraer a este cliente al no exigirle un contrato si otros proveedores sí se lo están exigiendo. Sin embargo, si otros proveedores tampoco están exigiendo contratos, parece que hay pocas más cosas que un proveedor pueda usar para tener ventaja sobre sus competidores. Aparentemente, este cliente será conquistado por la compañía que le ofrezca el precio fijo más bajo.

El análisis expuesto relativo a estos tres clientes ilustra el tipo de información que se puede obtener mediante el condicionamiento respecto a la elección de los clientes, y cómo la información puede traducirse fácilmente en una segmentación del mercado y en la identificación de oportunidades de marketing rentables.

### 11.6.3 Probabilidad condicionada para la última elección

Recordemos que la última situación de elección que afronta cada cliente no se incluyó en la estimación. Por lo tanto, se puede considerar una nueva situación de elección y se puede usar para evaluar el efecto de condicionar sobre decisiones pasadas. Para ello, identificamos qué alternativa escogió cada cliente en la nueva situación de elección y calculamos la probabilidad de elección de esa alternativa. La probabilidad se calculó en primer lugar sin condicionar en las elecciones anteriores. Este cálculo utiliza la fórmula del modelo logit mixto (11.5) empleando la distribución en la población de  $\beta_n$  y las estimaciones puntuales de los parámetros de la población. El promedio de esta probabilidad no condicionada entre los clientes es 0.353. A continuación se calculó la probabilidad condicionada a las elecciones anteriores. Se utilizaron cuatro maneras diferentes de calcular esta probabilidad:

1. Basada en la fórmula (11.6), utilizando las estimaciones puntuales de los parámetros de la población.
2. Basada en la fórmula (11.6), junto con el procedimiento descrito en la sección 11.3 que tiene en cuenta la varianza muestral de las estimaciones de los parámetros de la población.
3. (y 4). Empleando la fórmula logit

$$\frac{e^{\beta'_n x_{niT+1}}}{\sum_j e^{\beta'_n x_{njT+1}}},$$

usando la media condicionada  $\bar{\beta}_n$  como  $\beta_n$ . Este método es equivalente a utilizar la  $\bar{\beta}_n$  del cliente como si fuera una estimación de los verdaderos coeficientes del cliente,  $\beta_n$ . Las dos versiones (3 y 4) difieren en si  $\bar{\beta}_n$  se calcula sobre la base de la estimación puntual de los parámetros de la población (método 3) o se tiene en cuenta la distribución en la muestra de estos parámetros (método 4).



Los resultados se dan en la tabla 11.6 para el modelo 2. El resultado más destacado es que el condicionamiento sobre las elecciones anteriores de cada cliente mejora las previsiones para la última situación de elección de forma considerable. La probabilidad promedio de la alternativa elegida se incrementa de 0.35 sin condicionar a más de 0.50 al condicionar. Para cerca de tres cuartas partes de los 361 clientes en la muestra, la predicción de su última situación de elección es mejor al condicionar que al no hacerlo, incrementándose la probabilidad promedio de elección en más de 0.25. Para el resto de clientes, el condicionamiento hace que la predicción en las últimas situaciones de elección sea menos precisa, disminuyendo la probabilidad media de elección para estos clientes.

*Tabla 11.6. Probabilidad de la alternativa elegida en la última situación de elección*

	Método 1	Método 2	Método 3	Método 4
Probabilidad promedio	0.5213	0.5041	0.5565	0.5487
Número de clientes cuya probabilidad aumenta al condicionar	266	260	268	264
Incremento promedio en la probabilidad para clientes con incremento	0.2725	0.2576	0.3240	0.3204
Número de clientes cuya probabilidad disminuye al condicionar	95	101	93	97
Disminución promedio en la probabilidad para clientes con disminución	0.1235	0.1182	0.1436	0.1391

Hay varias razones por las que la probabilidad predicha después de condicionar no siempre es mayor. En primer lugar, los experimentos de elección se construyeron de manera que cada situación era bastante diferente del resto de situaciones, con el fin de obtener la mayor variación posible. Si la última situación implica nuevas comparaciones, las opciones anteriores no serán de utilidad y de hecho pueden ser perjudiciales para la predicción de la última elección. Una prueba más apropiada podría ser el diseño de una serie de situaciones de elección que obtenga información sobre comparaciones relevantes y luego diseñar una situación extra que esté dentro de la gama de comparaciones empleadas.

En segundo lugar, no hemos incluido en nuestro modelo todos los atributos de las alternativas que se presentaron a los clientes. En particular, hemos omitido los atributos que no entraron de manera significativa en la estimación de los parámetros de la población. Algunos clientes podrían responder a estos atributos omitidos, a pesar de que son insignificantes para la población en su conjunto. En la medida en que la última situación de elección implica comparaciones relativas a estos atributos, las distribuciones de preferencias condicionadas serían engañosas, dado que las preferencias relevantes quedarían excluidas. Esta explicación sugiere que si un modelo logit mixto va a ser usado para la obtención de densidades condicionadas para cada cliente, el investigador quizá debería incluir atributos que podrían ser importantes para algunos individuos a pesar de que son insignificantes para la población como un todo.

En tercer lugar, independientemente de cómo se ha diseñado la encuesta y el modelo, algunos clientes podrían responder a situaciones de elección de forma quijotesca, de manera que las preferencias evidenciadas en las elecciones anteriores no sean aplicadas por el cliente en la última situación de elección. Por último, factores aleatorios pueden hacer que la probabilidad de algunos clientes caiga al condicionar, incluso cuando las tres primeras razones no lo hagan.

Aunque al menos una de estas razones pudiera estar contribuyendo a reducir la probabilidad de elección para algunos de los clientes de nuestra muestra, la ganancia en la exactitud de predicción para los clientes con un aumento de la probabilidad después del condicionamiento, es más del doble que la

pérdida de precisión para aquellos con una disminución de la probabilidad, y el número de clientes con una ganancia es casi tres veces mayor que el número de cliente con una pérdida.

El tercer método (el más fácil), que se limita a calcular la fórmula logit estándar utilizando la  $\bar{\beta}_n$  de los clientes sobre la base de la estimación puntual de los parámetros de la población, da la mayor probabilidad. Este procedimiento no permite contemplar la distribución de  $\beta_n$  en torno a  $\bar{\beta}_n$  o la distribución muestral de  $\hat{\theta}$ . Permitir cualquiera de estas varianzas reduce la probabilidad media: usar la distribución condicionada de  $\beta_n$  en lugar de usar sólo la media  $\bar{\beta}_n$  (métodos 1 y 2 en comparación con los métodos 3 y 4, respectivamente) reduce la probabilidad media, y permitir la distribución muestral de  $\hat{\theta}$  en lugar de la estimación puntual (métodos 2 y 4 en comparación con los métodos 1 y 3, respectivamente) también reduce la probabilidad media. Este resultado no significa que el método 3, que incorpora la menor varianza, sea superior a los otros. Los métodos 3 y 4 son consistentes sólo si el número de situaciones de elección es capaz de incrementarse sin límite, de manera que  $\bar{\beta}_n$  pueda ser considerado como un estimador de  $\beta_n$ . Con  $T$  fijo, los métodos 1 y 2 son más adecuados, ya que incorporan toda la densidad condicionada.

### 11.7 Exposición

En este capítulo se muestra cómo la distribución de los coeficientes condicionada a las elecciones observadas del cliente se obtiene a partir de la distribución de los coeficientes de la población. Si bien estas distribuciones condicionadas pueden ser útiles de varias formas, es importante reconocer las limitaciones del concepto. En primer lugar, el uso de distribuciones condicionadas para hacer predicciones se limita a aquellos clientes cuyas elecciones anteriores han sido observadas. En segundo lugar, mientras que la distribución condicionada de cada cliente puede utilizarse en el análisis de conglomerados (*cluster analysis*) así como para otros fines de identificación, el investigador a menudo quiere relacionar preferencias con características demográficas observables de los clientes. Sin embargo, estos datos demográficos observables de los clientes podrían ser introducidos directamente en el modelo, de modo que los parámetros de la población varíen con las características observadas de los clientes en la población. De hecho, introducir demográficos en el modelo es más directo y más accesible para hacer pruebas de hipótesis que estimar un modelo sin esas características, calcular la distribución condicionada para cada cliente y luego hacer el análisis de conglomerados y otros análisis usando los momentos estadísticos de las distribuciones condicionadas.

Teniendo en cuenta estas cuestiones, hay tres razones principales por las que un investigador podría beneficiarse del cálculo de las distribuciones condicionadas de los clientes. En primer lugar, cada vez resulta más simple disponer de información sobre elecciones pasadas. Algunos ejemplos de esta disponibilidad incluyen los datos que provienen de un lector de códigos de barras de clientes que compran en tiendas de comestibles usando una tarjeta de fidelización, los programas para viajeros frecuentes que ofrecen las aerolíneas y las compras de particulares a través de Internet. En estas situaciones, condicionar sobre elecciones pasadas permite una comercialización específica eficaz y el desarrollo de nuevos productos y servicios que respondan a las preferencias reveladas de los subgrupos de clientes.

En segundo lugar, las características demográficas que diferencian a los clientes con diferentes preferencias podrían ser más evidentes a través de un análisis de conglomerados aplicado sobre distribuciones condicionadas, que a través de pruebas de especificación en el modelo mismo. El análisis de conglomerados tiene su propia forma de identificar patrones, que puede ser en algunos casos más eficaz que las pruebas de especificación dentro de un modelo de elección discreta.

En tercer lugar, el examen de las distribuciones condicionadas de los clientes a menudo puede identificar patrones que no pueden relacionarse con características observadas de esos clientes, pero que sin

embargo es útil conocerlos. Por ejemplo, saber que un producto o una campaña de marketing serán de interés para una parte de la población debido a sus preferencias particulares suele ser suficiente, sin necesidad de identificar a ese grupo de personas sobre la base de sus características demográficas. Las densidades condicionadas pueden facilitar enormemente los análisis que tienen estos objetivos.