Эконометрический ликбез: ограниченные зависимые переменные

Мультиномиальные модели дискретного выбора*

Золт Шандор†

Университет Гронингена, Гронинген, Нидерланды

В настоящем эссе описаны основные свойства хорошо известных мультиномиальных моделей дискретного выбора: стандартной логит-модели, смешанной логит-модели и пробит-модели.

1 Введение

В эссе описываются некоторые модели дискретного выбора, применяемые для оценивания параметров спроса на товары, приобретаемые в дискретных количествах. В этих моделях различные товары рынка считаются различными выборными альтернативами. В интересующих нас моделях предполагается, что в течение заданного периода времени потребитель, считающийся принимающие решения лицом, приобретает либо одну, либо ноль единиц товара. Набор всех альтернатив, из которых выбирает потребитель, считается конечным и содержащим все товары, доступные на рынке. Обозначим альтернативу как j, а набор всех альтернатив как $\{1, 2, ..., J\}$. В модели дискретного выбора потребитель i, выбирающий альтернативу j, приобретает случайную полезность u_{ij}, которая напрямую исследователем не наблюдается. Во многих ситуациях, однако, вероятность того, что выбрана альтернатива j, наблюдается с небольшой выборочной ошибкой в виде доли рынка. Следовательно, обычно предполагается, что i максимизирует свою полезность, выбирая j таким образом, что u_{ij} наибольшая из всех u_{ir}, $r = 1, ..., J$. Вероятность того, что i выбирает j, можно вычислить, если мы знаем распределения, лежащие в основе функции полезности. В настоящем эссе описаны основные свойства хорошо известных мультиномиальных моделей дискретного выбора: стандартной логит-модели, смешанной логит-модели и пробит-модели.

Изначальным источником логита и пробита было приложение вероятностных моделей к биологическим опытам бинарного выбора в первой половине прошлого века. Первые приложения в экономике использовали биномиальную пробит-модель (например, Farrell, 1954), которая, по-видимому, является победителем в споре между биологами по поводу способности логита и пробита моделировать бинарный выбор. Theil (1969) обобщил биномиальную логит-модель до мультиномиальной логит-модели, открыв дорогу дальнейшим улучшениям и приложениям. В начале 70-х McFadden с соавторами, которые изучали некоторые исследовательские проблемы транспортировки, обобщили логит-модель в нескольких направлениях и сделали ее научно признанной, предоставив теоретическую основу в теории полезности дискретного выбора (например, McFadden, 1973; McFadden & Reid, 1975; McFadden, 1977). Более детальные исторические данные содержатся в Cramer (1991, стр. 39–42) и Anderson, de Palma & Thisse (1992, глава 2).

*Перевод М. Кузина и С. Анатольева. Цитировать как: Шандор, Золт (2009). «Мультиномиальные модели дискретного выбора», Квантиль, №7, стр. 9–19. Citation: Sándor, Zsolt (2009). "Multinomial discrete choice models," Quantile, No.7, pp. 9–19.

†Адрес: Department of Economics and Econometrics, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. Электронная почта: *Z.Sandor@rug.nl*

2 Стандартная логит-модель

В этом разделе мы кратко представляем стандартную логит-модель, обсуждаем ее главные теоретические преимущества и практические недостатки. Затем мы выведем некоторые основные формулы, связанные с асимптотическими свойствами оценки максимального правдоподобия модели.

2.1 Спецификация модели

В стандартной логит-модели полезность является линейной функцией свойств альтернативы:

$$u_{ij} = x_{ij}'\beta + \varepsilon_{ij}, \tag{1}$$

где x_{ij} – $K \times 1$ вектор, содержащий характеристики потребителя i и альтернативы j, β – $K \times 1$ вектор параметров, а переменные ε_{ij}, $j = 1, ..., J$, предполагаются случайными и имеющими независимые стандартные распределения экстремальных значений типа I, кумулятивная функция распределения которого равна

$$F\left(\varepsilon_{ij}\right) = \exp\left(-\exp\left(-\varepsilon_{ij}\right)\right), \tag{2}$$

а функция плотности равна

$$f\left(\varepsilon_{ij}\right) = \exp\left(-\varepsilon_{ij}\right) \cdot \exp\left(-\exp\left(-\varepsilon_{ij}\right)\right). \tag{3}$$

В силу принципа максимизации полезности, вероятность того, что i выбирает альтернативу j, есть

$$s_{ij} = \Pr\left\{u_{ij} \geq u_{ir},\ \text{для всех } r = 1, ..., J\right\}.$$

Используя выражение для полезности (1), эта вероятность равна

$$s_{ij} = \Pr\left\{\varepsilon_{ir} \leq \varepsilon_{ij} + x_{ij}'\beta - x_{ir}'\beta,\ \text{для всех } r = 1, ..., J\right\}.$$

Вероятности такого типа обычно вычисляют как J-мерный интеграл, используя совместное распределение случайных элементов:

$$s_{ij} = \int \cdots \int_{\substack{\varepsilon_{ir} \leq \varepsilon_{ij} + x_{ij}'\beta - x_{ir}'\beta \\ \text{для всех } r=1,...,J}} f\left(\varepsilon_{i1}, ..., \varepsilon_{iJ}\right) \mathrm{d}\varepsilon_{i1} ... \mathrm{d}\varepsilon_{iJ}.$$

Так как все ε независимы, имеем

$$s_{ij} = \int \cdots \int_{\substack{\varepsilon_{ir} \leq \varepsilon_{ij} + x_{ij}'\beta - x_{ir}'\beta \\ \text{для всех } r=1,...,J}} f\left(\varepsilon_{i1}\right) \cdots f\left(\varepsilon_{iJ}\right) \mathrm{d}\varepsilon_{i1} ... \mathrm{d}\varepsilon_{iJ},$$

что равно

$$\int_{\mathbb{R}} f\left(\varepsilon_{ij}\right) \int \cdots \int_{\substack{\varepsilon_{ir} \leq \varepsilon_{ij} + x_{ij}'\beta - x_{ir}'\beta \\ r \neq j}} \prod_{r \neq j} f\left(\varepsilon_{ir}\right) \mathrm{d}\varepsilon_{i1} ... \mathrm{d}\varepsilon_{iJ}.$$

Наиболее крупная, $(J-1)$-мерная часть этого интеграла равна произведению $J-1$ одномерных интегралов плотностей, так что мы можем использовать кумулятивные функции распределения:

$$s_{ij} = \int_{\mathbb{R}} f\left(\varepsilon_{ij}\right) \cdot \prod_{r \neq j} F\left(\varepsilon_{ij} + x_{ij}'\beta - x_{ir}'\beta\right) \mathrm{d}\varepsilon_{ij}.$$

Используя выражения для этих функций (2) и (3) , получаем

$$s_{ij} = \int_{\mathbb{R}} \exp\left(-\varepsilon_{ij}\right) \cdot \exp\left(-\sum_{r=1}^{J} \exp\left(-\left(\varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta\right)\right)\right) \mathrm{d}\varepsilon_{ij}.$$

Этот интеграл имеет явный вид и может быть вычислен следующим способом:

$$s_{ij} = \frac{1}{\sum_{r=1}^{J} \exp\left(-(x'_{ij}\beta - x'_{ir}\beta)\right)} \exp\left(-\sum_{r=1}^{J} \exp\left(-\left(\varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta\right)\right)\right)\Bigg|_{\varepsilon_{ij}=-\infty}^{\infty},$$

что в итоге дает

$$s_{ij} = \frac{\exp(x'_{ij}\beta)}{\sum_{r=1}^{J} \exp\left(x'_{ir}\beta\right)}. \tag{4}$$

Благодаря этой привлекательной явной форме вероятности в (4) стандартная логит-модель и популярна в тех дисциплинах, в которых используются модели выбора.

В следующих разделах мы применим стандартную логит-модель в ситуациях, в которых характеристики варьируются по продуктам и не варьируются по потребителям, то есть когда $x_{ij} = x_j$ для всех $j = 1, ..., J$. Благодаря этому вероятность того, что потребитель i выбирает продукт j, равна

$$s_j = \frac{\exp(x'_j\beta)}{\sum_{r=1}^{J} \exp\left(x'_r\beta\right)} \tag{5}$$

для всех потребителей. Следовательно, это вероятность того, что продукт j приобретается на рынке. Эта величина на рынке с большим числом потребителей равно доле рынка продукта j.

Хотя выражение для вероятности того, что выбрана такая-то альтернатива, выглядит просто и потому доступно для практиков в вычислительном смысле, стандартная логит-модель также имеет серьезный недостаток. Этот недостаток возникает из-за простоты модели для полезности, а именно, из-за того, что полезности всех потребителей зависят от скалярной функции характеристик, $x'_j\beta$, а не от всего вектора характеристик x_j, поэтому отношение вероятностей двух альтернатив j и q не зависит от наличия и свойств любых других альтернатив, так как

$$\frac{s_j}{s_q} = \frac{\exp(x'_j\beta)}{\exp\left(x'_q\beta\right)}. \tag{6}$$

Это свойство известно как независимость от посторонних альтернатив. Это непривлекательное свойство, так как если мы добавляем новую альтернативу к набору всех альтернатив, которые являются близкими заменителями для j, но не для q, мы ожидаем, что s_j снизится значительно сильнее, чем s_q. Поэтому отношение двух вероятностей также должно уменьшиться.

Мы проиллюстрируем эту проблему на простом примере. Предположим, что на рынке есть два товара со следующими свойствами $x_1 = (1, 1, 1)'$ и $x_2 = (.5, 1.5, 1)'$, и это единственные альтернативы для выбора. Пусть $\beta = (1, 1, 1)'$. Тогда $s_1 = s_2 = \frac{1}{2}$. Теперь введем третий продукт со свойствами $x_3 = (1, 1, 1)'$. В этой новой ситуации $s_1 = s_2 = s_3 = \frac{1}{3}$, то есть стандартная логит-модель предсказывает, что потребители будут замещать продукты 1 и 2 в одинаковой степени. Это неверно, так как приблизительно половина потребителей предпочитает продукт 1, который является близким заменителем товару 2, так что часть потребителей, предпочитающих 1, должны переключиться на 3, а те, кто предпочитает 2, должны остаться с 2.

2.2 Логарифмическое правдоподобие и свойства ее первой производной

Так как далее мы используем результаты, связанные с оцениванием стандартной логит-модели, здесь мы приведем некоторые из этих результатов. Оценивание параметров стандартной логит-модели основано на идее о том, что мы находимся в ситуации мультиномиального выбора, когда вероятность s_j того, что выбрана альтернатива j, известна. Тогда если мы наблюдаем большое число выборов, частота альтернативы j, то есть число раз, когда j выбрана, деленное на общее число выборов, должно равняться s_j. Обозначим частоту альтернативы j за f_j. Пусть n – общее число выборов, то есть число потребителей. Тогда $n \cdot f \equiv n \cdot (f_1, ..., f_J)'$ имеет мультиномиальное распределение с параметрами $s_1, ..., s_J$. Поэтому вероятность того, что альтернативы $1, ..., J$ встречаются с частотами $f_1, ..., f_J$, соответственно, на самом деле равна функции правдоподобия, соответствующей стандартной логит-модели:

$$L = C \cdot \left(s_1^{f_1} \cdots s_J^{f_J} \right)^n,$$

где C – коэффициент, соответствующий мультиномиальному распределению. Так как этот коэффициент не зависит от интересующих нас параметров, в дальнейшем мы игнорируем его, когда записываем логарифмическую функцию правдоподобия, которая равна

$$\ln L = n \cdot \sum_{j=1}^{J} f_j \ln s_j. \tag{7}$$

Оценивание параметров можно выполнить максимизацией логарифмической функции правдоподобия. Обозначим за S диагональную матрицу с диагональю $s = (s_1, ..., s_J)'$. Можно записать $\ln L = n \cdot f' \ln s$. Следовательно, производная первого порядка логарифмической функции правдоподобия может быть записана как

$$\frac{\partial \ln L}{\partial \beta} = n \cdot \left(f' \frac{\partial \ln s}{\partial \beta'} \right)' = n \cdot \left(\frac{\partial s}{\partial \beta'} \right)' S^{-1} f,$$

так как

$$\frac{\partial \ln s}{\partial \beta'} = S^{-1} \frac{\partial s}{\partial \beta'}.$$

Для того чтобы вывести $\dfrac{\partial s}{\partial \beta'}$, заметим, что

$$\frac{\partial s_j}{\partial \beta'} = x_j' s_j - s' X s_j,$$

где $X = (x_1 \ ... \ x_J)'$, что приводит к

$$\frac{\partial s}{\partial \beta'} = \left(S - ss' \right) X. \tag{8}$$

Эту формулу можно использовать для того чтобы получить первые производные логарифмической функции правдоподобия:

$$\frac{\partial \ln L}{\partial \beta} = n \cdot X' \left(S - ss' \right) S^{-1} f.$$

Теперь можно вычислить асимптотическую информационную матрицу:

$$\mathcal{I}\left(\beta | X\right) = \mathbb{E}\left[\frac{\partial \ln L}{\partial \beta} \frac{\partial \ln L}{\partial \beta'} \right] = n^2 \cdot X' \left(S - ss' \right) S^{-1} \mathbb{E}\left[f f' \right] S^{-1} \left(S - ss' \right) X.$$

Мультиномиальное распределение имеет свойство $\mathrm{E}[ff'] = \frac{1}{n}S + \left(1 - \frac{1}{n}\right)ss'$. Подставляя это выражение в предыдущую формулу, после некоторых вычислений получаем

$$\mathcal{I}\left(\beta|X\right) = n \cdot X'\left(S - ss'\right)X.$$

Эта формула использовалась напрямую для построения модели с улучшенной эффективностью для экспериментов выбора (см. например, Sandor & Wedel, 2001).

3 Смешанная логит-модель

Смешанная логит-модель не удовлетворяет условию независимости от посторонних альтернатив. Это происходит из-за того, что в спецификации полезности коэффициенты предполагаются случайными. Таким образом, потребители имеют различные коэффициенты в функциях полезности. Такая спецификация имеет следующее привлекательное свойство: она параметризует отношение потребителей к характеристикам альтернатив через вариацию случайных параметров. Как в стандартной логит-модели, первые экономические приложения смешанной логит-модели, по-видимому, встретились в сфере транспортных исследований (Boyd & Mellman, 1980; Cardell & Dunbar, 1980). Мы отсылаем читателей к более детальным обзорам смешанных логит-моделей McFadden & Train (2000) и Brownstone & Train (1999). В данном разделе мы представим модель и основные асимптотические свойства оценки максимального правдоподобия.

3.1 Спецификация модели

В структуре выбора, аналогичной случаю стандартной логит-модели, полезность потребителя i, выбирающего альтернативу j, определяется как

$$u_{ij} = x_j'\beta_i + \varepsilon_{ij}, \tag{9}$$

где единственное отличие от (1) заключается в том, что здесь β_i – вектор параметров, характерных для потребителя i. Мы предполагаем, что $\beta_i \sim N(\beta, \Sigma)$, $i = 1, ..., n$, независимые и одинаково распределенные случайные величины, где β – $K \times 1$ вектор параметров, а Σ – симметричная положительно определенная матрица параметров. Ошибки ε_{ij} предполагаются независимыми от β_i.

Для удобства вычислений запишем случайный вектор параметров в следующем виде:

$$\beta_i = \beta + \Lambda v_i, \tag{10}$$

где Λ – некая матрица, такая, что $\Sigma = \Lambda\Lambda'$, а v_i – $K \times 1$ случайный вектор, имеющий стандартное нормальное распределение.

Если бы мы знали v_i, мы могли бы использовать формулу (5), поскольку это была бы стандартная логит-модель. В этом случае вероятность того, что альтернатива j выбрана, была бы

$$\Pr\{j|v_i\} = \frac{\exp\left(x_j'\left(\beta + \Lambda v_i\right)\right)}{\sum_{r=1}^{J}\exp\left(x_r'\left(\beta + \Lambda v_i\right)\right)}. \tag{11}$$

Безусловная вероятность того, что альтернатива j выбрана, определяется формулой

$$s_j = \mathbb{E}\left[\Pr\{j|v_i\}\right] = \int_{\mathbb{R}^K} \frac{\exp\left(x_j'\left(\beta + \Lambda v_i\right)\right)}{\sum_{r=1}^{J}\exp\left(x_r'\left(\beta + \Lambda v_i\right)\right)}\phi\left(v_i\right)\mathrm{d}v_i, \tag{12}$$

где $\phi(v_i)$ – функция плотности v_i. Заметим, что поскольку (6) в данном случае не выполняется, то смешанная логит-модель не удовлетворяет свойству независимости от посторонних альтернатив.

Выражение для вероятности в (12) представляет собой интеграл, который в общем случае нельзя вычислить аналитически. Обычно он оценивается с помощью симуляций Монте-Карло или квази-Монте-Карло. Детальное обсуждение технологии такого оценивания выходит за рамки данного эссе, поэтому мы лишь упомянем здесь ее суть. Извлекаем несколько значений случайного вектора v_i, вычисляем функцию под интегралом при каждом его значении, и находим среднее по этим значениям. Полученное среднее и есть (квази-) оценка интеграла методом Монте-Карло.

Чтобы проиллюстрировать, как в смешанной логит-модели моделируется отношение потребителей к различным характеристикам, рассмотрим более простой вид ковариационной матрицы, а именно, когда она диагональна:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_K^2 \end{pmatrix}.$$

В этом случае можно взять

$$\Lambda = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_K \end{pmatrix},$$

так что можно записать полезность в (9) с использованием (10) как

$$u_{ij} = x_j'\beta + \sum_{k=1}^{K} \sigma_k v_{ik} x_{jk} + \varepsilon_{ij}.$$

Записав полезность в таком виде, мы можем проинтерпретировать эффект случайной части параметров. Каждый потребитель i сопоставляется с вектором v_i, элементы которого представляют собой оценку потребителем характеристик. Если предпочтения i таковы, что для него более ценны альтернативы с большими значениями какой-то характеристики, то у такого потребителя будет большое положительное значение v_i, соответствующее этой характеристике. Например, в случае автомобилей, если потребитель i предпочитает большие машины, и «размер» – это характеристика h в этой модели, то этот потребитель будет иметь большое положительное v_{ih}.

Теперь можно объяснить преимущества смешанной логит-модели перед обычной тем, что схемы замен, порождаемые смешанной логит-моделью, подвергаются меньшим ограничениям. Продолжая пример с машинами, если на рынке появляется маленькая машина, то потребитель, о котором шла речь выше, не будет сильно стремиться к переключению на новую маленькую машину. Происходит это из-за того, что для этого потребителя значение $\sigma_h v_{ih}$ несоразмерно больше, чем соответствующие величины для других характеристик, и поэтому член $\sigma_h v_{ih} x_{jh}$ более чувствителен к вариациям в x_{jh}, чем в случае с другими характеристиками. Поэтому полезность потребителя i высока для больших значений x_{jh} и низка для малых.

Описанное выше подразумевает, что чем больше изменения параметров отношения потребителей, σ_k, тем менее вероятно, что потребитель, предпочитающий определенную альтернативу, будет переключаться на несхожую альтернативу. Проиллюстрируем это примером из конца раздела 2.1. В дополнение к случаю стандартной логит-модели здесь мы возьмем $(\sigma_1, \sigma_2, \sigma_3) = \lambda(1,1,1)$. Когда только товары 1 и 2 находятся на рынке, их вероятности $s_1 = s_2 = \frac{1}{2}$, независимо от значения λ. Когда товар 3 появляется на рынке, s_2 становится

0.35 при $\lambda = 1$, 0.42 при $\lambda = 5$, 0.47 при $\lambda = 10$ и 0.49 при $\lambda = 30$. Таким образом, для больших вариаций нет замещения товаром 3 товара 2, в то время как для малых вариаций замещение очень похоже на то, что возникало в стандартной логит-модели.

3.2 Логарифмическое правдоподобие и свойства ее первой производной

В этом разделе мы выведем некоторые полезные формулы для смешанной логит-модели. Обозначим условную вероятность в (11) за $\pi_j(v)$. Тогда

$$s_j = \int_{\mathbb{R}^K} \pi_j(v)\,\phi(v)\,\mathrm{d}v.$$

Модель выбора в данном случае можно сформулировать похожим образом, как для стандартной логит-модели. Тогда функция логарифмического правдоподобия в этой модели равна

$$\ln L = n \cdot \sum_{j=1}^{J} f_j \ln s_j,$$

где f_j вновь обозначает наблюдаемую частоту товара j.

Чтобы сократить обозначения, мы пишем интегралы типа $\int_{\mathbb{R}^K}(\cdot)\phi(v)\mathrm{d}v$ как $\int(\cdot)\mathrm{d}\Phi$ (например, $\int_{\mathbb{R}^K}\pi_j(v)\,\phi(v)\mathrm{d}v \equiv \int \pi_j \mathrm{d}\Phi$). При оценивании методом максимального правдоподобия мы находим первые производные функции логарифмического правдоподобия. Так как нам нужны эти результаты для случая, когда ковариационная матрица случайных коэффициентов Σ диагональна, мы выводим формулы только для этого случая. Для удобства вычислений мы изменили обозначение случайной части коэффициентов с Λv на $V\sigma$, где V – диагональная матрица с диагональю v, а $\sigma = (\sigma_1, ..., \sigma_K)'$. Аналогично случаю стандартной логит-модели,

$$\frac{\partial L}{\partial \beta} = n \cdot \left(\frac{\partial s}{\partial \beta'}\right)' S^{-1} f \quad \text{и} \quad \frac{\partial L}{\partial \sigma} = n \cdot \left(\frac{\partial s}{\partial \sigma'}\right)' S^{-1} f. \tag{13}$$

Далее нам потребуются следующие формулы:

$$\frac{\partial s}{\partial \beta'} = \int \frac{\partial \pi}{\partial \beta'}\mathrm{d}\Phi \quad \text{и} \quad \frac{\partial s}{\partial \sigma'} = \int \frac{\partial \pi}{\partial \sigma'}\mathrm{d}\Phi,$$

где $\pi = (\pi_1, ..., \pi_J)'$. Аналогично, как для стандартной логит-модели, получаем:

$$\frac{\partial \pi}{\partial \beta'} = \left(\Pi - \pi\pi'\right) X \quad \text{и} \quad \frac{\partial \pi}{\partial \sigma'} = \left(\Pi - \pi\pi'\right) XV,$$

где Π – диагональная матрица с диагональю π. Отсюда

$$\frac{\partial s}{\partial \beta'} = \int \left(\Pi - \pi\pi'\right) X \mathrm{d}\Phi \quad \text{и} \quad \frac{\partial s}{\partial \sigma'} = \int \left(\Pi - \pi\pi'\right) XV \mathrm{d}\Phi. \tag{14}$$

Подставляя эти равенства в (13), получаем:

$$\frac{\partial L}{\partial \beta} = n \cdot \int X' \left(\Pi - \pi\pi'\right) \mathrm{d}\Phi \cdot S^{-1} f \tag{15}$$

и

$$\frac{\partial L}{\partial \sigma} = n \cdot \int V X' \left(\Pi - \pi\pi'\right) \mathrm{d}\Phi \cdot S^{-1} f. \tag{16}$$

Асимптотическая информационная матрица определяется как

$$\mathcal{I}(\beta, \sigma | X) = \begin{bmatrix} \mathbb{E}\left[\frac{\partial L}{\partial \beta}\frac{\partial L}{\partial \beta'}\right] & \mathbb{E}\left[\frac{\partial L}{\partial \beta}\frac{\partial L}{\partial \sigma'}\right] \\ \mathbb{E}\left[\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \beta'}\right] & \mathbb{E}\left[\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \sigma'}\right] \end{bmatrix}.$$

Для упрощения этого выражения используем тот факт, что по предположению f имеет мультиномиальное распределение с $\mathbb{E}[f] = s$. Тогда f обладает свойством $\mathbb{E}[ff'] = \frac{1}{n}S + \left(1 - \frac{1}{n}\right) ss'$. Используя этот факт и формулы (15) и (16), находим:

$$\mathcal{I}\left(\beta, \sigma | X\right) = n \cdot \left[\begin{array}{cc} M'S^{-1}M & M'S^{-1}Q \\ Q'S^{-1}M & QS^{-1}Q \end{array} \right], \tag{17}$$

где

$$M = \int \left(\Pi - \pi\pi'\right) X \mathrm{d}\Phi \text{ и } Q = \int \left(\Pi - \pi\pi'\right) XV \mathrm{d}\Phi.$$

Формула (17) использовалась напрямую для построения экспериментов выбора для смешанной логит-модели (см. например, Sandor & Wedel, 2002).

4 Пробит

Пробит – это модель, похожая на смешанную логит-модель, в которой все случайные элементы, коэффициенты и ошибки по предположению имеют нормальное распределение. Одно из первых применений мультиномиальной пробит-модели в эконометрическом контексте встретилось в работе McFadden (1976). В этом разделе мы кратко представим модель и покажем способ оценивания вероятностей выбора, известный как GHK-симулятор (по первым буквам Geweke, Hajivassiliou и Keane, разработчиков симулятора для эконометрического контекста в начале 90-х). В отличие от стандартной и смешанной логит-моделей, для пробит-модели мы не приводим функцию логарифмического правдоподобия и свойства ее первых производных. Их краткое описание содержится в Wansbeek & Webel (2001).

4.1 Спецификация модели

Предположим, что полезности, соответствующие J продуктам для потребителя i, равны

$$u_{ij} = x_j'\beta_i + \varepsilon_{ij},$$

где, как и для смешанной логит-модели, $\beta_i \sim N(\beta, \Sigma)$, $i = 1, ..., n$, независимо и одинаково распределены, β – $K \times 1$ вектор параметров, а Σ – симметричная положительно определенная матрица параметров. Ошибка ε_{ij} здесь предполагается нормально распределенной без требования независимости от β_i. Благодаря привлекательному свойству нормального распределения, а именно, тому, что сумма двух нормально распределенных случайных величин нормально распределена, можно выбрать подходящие переменные μ и Ω таким образом, что вектор полезностей можно представить как

$$u_i = \mu + \Gamma e_i \quad \sim N\left(\mu, \Omega\right), \tag{18}$$

где $\mu : J \times 1$, $\Gamma : J \times J$, такая что $\Gamma\Gamma' = \Omega$, и $e_i \sim N\left(0, I_J\right)$. Если случайные члены e_i независимы, то можно опустить индекс i, и каждому потребителю сопоставить реализацию случайного вектора e. Отметим, что если в полезности в смешанной логит-модели (9) мы заменим предположение о распределении экстремальных значений остатков нормальным распределением, то получим частный случай (18). Причина, по которой последняя спецификация более общая, заключается в том, что в данном случае не предполагается независимости между остатками и случайными коэффициентами.

Из J продуктов выбран один с наибольшей полезностью. Вероятность того, что был выбран продукт j, равна

$$s_j = \Pr\left\{u_j \geq u_r, \forall r\right\} = \Pr\left\{V_j u \leq 0\right\} = \Pr\left\{V_j\mu + V_j\Gamma e \leq 0\right\}$$

с

$$V_j = \begin{bmatrix} 1 & 0 & \dots & -1 & \dots & 0 \\ 0 & 1 & \dots & -1 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & -1 & \dots & 1 \end{bmatrix},$$

где столбец, состоящий из -1, стоит на j-ом месте. Обозначим $T \equiv V_j\Gamma$ и выберем Γ так, что матрица T нижняя треугольная. Это возможно сделать, так как $TT' = V_j\Omega V_j'$, а эта матрица положительно определена. Тогда

$$s_j = \Pr\{Te \leq v\}, \text{ где } v \equiv -V_j\mu. \tag{19}$$

Если мы обозначим

$$T = \begin{bmatrix} t_{11} & 0 & \dots & 0 \\ t_{12} & t_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ t_{J1} & t_{J2} & \dots & t_{JJ} \end{bmatrix},$$

то

$$s_j = \Pr\left\{e_1 \leq \frac{v_1}{t_{11}}, e_2 \leq \frac{v_2 - t_{21}e_1}{t_{22}}, \dots, \; e_J \leq \frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right\}.$$

Пусть

$$D = \left\{(e_1, ..., e_J) \in \mathbb{R}^J : e_1 \leq \frac{v_1}{t_{11}}, ..., e_J \leq \frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right\}.$$

Тогда формула для вероятности того, что выбрана альтернатива j, выглядит как

$$s_j = \int_D \phi(e_1) ... \phi(e_J) \, \mathrm{d}e_1 ... \mathrm{d}e_J, \tag{20}$$

где ϕ – плотность стандартного нормального распределения.

4.2 Оценивание вероятностей

Чтобы получить более удобные в вычислительном смысле формулы вероятности, заметим, что e имеют усеченное нормальное распределение. Если $e \leq b$ и $e \sim N(0,1)$, тогда можно получить e быстрее, извлекая u равномерно распределенным на $[0,1]$ и беря $e = \Phi^{-1}(u \cdot \Phi(b))$, где Φ – стандартная нормальная функция распределения. Оказывается, что основываясь на этой идее, можно трансформировать интеграл (20) в интеграл, аргументы которого принадлежат единичному гиперкубу.

Для того чтобы произвести трансформацию, можно воспользоваться идеей из предыдущего раздела в J-мерном случае. Это подразумевает трансформацию

$$\begin{aligned} e_1 &= \Phi^{-1}\left(u_1\Phi\left(\frac{v_1}{t_{11}}\right)\right) \equiv \psi_1(u_1, ..., u_J), \\ e_2 &= \Phi^{-1}\left(u_2\Phi\left(\frac{v_2 - t_{21}e_1(u_1)}{t_{22}}\right)\right) \equiv \psi_2(u_1, ..., u_J), \\ &\vdots \\ e_J &= \Phi^{-1}\left(u_J\Phi\left(\frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right)\right) \equiv \psi_3(u_1, ..., u_J), \end{aligned} \tag{21}$$

где $u_1, ..., u_J \in [0,1]$, и в последней формуле e_j рассматривается как функция от $u_1, ..., u_j$ для $j = 1, ..., J-1$. Тогда (20) становится

$$\int_{[0,1]^J} \phi\left(\psi_1\left(u_1, ..., u_J\right)\right) ... \phi\left(\psi_J\left(u_1, ..., u_J\right)\right) |J| \, \mathrm{d}u_1 ... \mathrm{d}u_J,$$

где $|J|$ – якобиан, то есть

$$|J| = \begin{vmatrix} \frac{\partial \psi_1}{\partial u_1} & \cdots & \frac{\partial \psi_J}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \psi_1}{\partial u_J} & \cdots & \frac{\partial \psi_J}{\partial u_J} \end{vmatrix} = \begin{vmatrix} \frac{1}{\phi(\psi_1(...))} \Phi\left(\frac{v1}{t_{11}}\right) & \cdots & \frac{\partial \psi_J}{\partial u_1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\phi(\psi_J(...))} \Phi\left(\frac{v_J - t_{J1} e_1 - ... - t_{JJ-1} e_{J-1}}{t_{JJ}}\right) \end{vmatrix},$$

что является определителем верхней треугольной матрицы, поэтому якобиан равен произведению диагональных элементов:

$$|J| = \frac{\Phi\left(\frac{v_1}{t_{11}}\right) \cdots \Phi\left(\frac{v_J - t_{J1} e_1 - ... - t_{JJ-1} e_{J-1}}{t_{JJ}}\right)}{\phi\left(\psi_1\left(...\right)\right) \cdots \phi\left(\psi_J\left(...\right)\right)}.$$

Используя полученные формулы, можно преобразовать (20) в

$$\int_{[0,1]^J} \Phi\left(\frac{v_1}{t_{11}}\right) \cdots \Phi\left(\frac{v_J - t_{J1} e_1 - ... - t_{JJ-1} e_{J-1}}{t_{JJ}}\right) \mathrm{d}u_1 ... \mathrm{d}u_J.$$

Тогда получаем формулу для вероятности

$$s_j = \Phi\left(\frac{v_1}{t_{11}}\right) \int_{[0,1]^{J-1}} \Phi\left(\frac{v_2 - t_{21} e_1}{t_{22}}\right) \cdots \Phi\left(\frac{v_J - t_{J1} e_1 - ... - t_{JJ-1} e_{J-1}}{t_{JJ}}\right) \mathrm{d}u_1 ... \mathrm{d}u_{J-1}.$$

Интеграл в этой формуле можно интерпретировать как ожидание от подынтегральной функции при равномерно распределенном случайном $(J-1)$-векторе. Этот интеграл можно оценить с помощью симуляций Монте-Карло, извлекая набор случайных векторов и вычисляя их среднее. Оценка, найденная таким способом, называется GHK-симулятором (Börsch-Supan & Hajivassiliou, 1993) или RIS-симулятором (от *recursive importance sampling*), основанным на усеченной нормальной плотности (Vijverberg, 1997). Vijverberg (1997) также обсуждает другие RIS-симуляторы. Ряд других типов симуляторов представлен в работе Hajivassiliou, McFadden & Ruud (1996).

Благодарности

Автор благодарит Тома Вансбеека за многочисленные полезные комментарии.

Список литературы

Anderson, S.P., A. de Palma & J-F. Thisse (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge: MIT Press.

Börsch-Supan, A. & V.A. Hajivassiliou (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* 58, 347–368.

Boyd, H.J. & R.E. Mellman (1980). The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis. *Transportation Research* 14, 367–378.

Brownstone, D. & K. Train (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89, 109–129.

Cardell, N.S. & F. Dunbar (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research* 14, 423–434.

Cramer, J.S. (1991). *The Logit Model: An Introduction for Economists.* London: Arnold.

Farrell, M.J. (1954). The demand for motorcars in the United States. *Journal of the Royal Statistical Society, A* 117, 171–200.

Hajivassiliou, V., D. McFadden & P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics* 72, 85–134.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. Глава в *Frontiers in Econometrics* под редакцией P. Zarembka. New York: Academic Press.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement* 5, 363–90.

McFadden, D. (1977). Econometric models of probabilistic choice. Стр. 171–260 в *Structural Analysis of Discrete Data with Econometric Applications* под редакцией C.F. Manski & D. McFadden. Cambridge: MIT Press.

McFadden, D. & F. Reid (1975). Aggregate travel demand forecasting from disaggregated behavioral models. *Transportation Research Record* 534, 24–37.

McFadden, D. & K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15, 447–470.

Sándor, Z. & M. Wedel (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38, 430–444.

Sándor, Z. & M. Wedel (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21, 55–75.

Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review* 10, 251–259.

Vijverberg, W.P.M. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics* 76, 281–307.

Wansbeek, T. & M. Wedel (2001). The structure of the multinomial probit Model. Working Paper, University of Groningen.

Multinomial discrete choice models

Zsolt Sándor

University of Groningen, Groningen, Netherlands

This essay briefly describes the main features of some well-known multinomial discrete choice models: the standard logit, the mixed logit, and the probit.