

COE 特別講義
ミクロ計量経済学講義
(Lectures on Micro-econometrics)

雨宮 健 (Takeshi Amemiya)
スタンフォード大学経済学科
(Department of Economics, Stanford University)

2005年6月

前書き (Preface)

計量経済学 (Econometrics) の分野では、しばらく前よりマイクロ計量経済学 (Micro-econometrics) と呼ばれているマイクロ・レベルでの計量分析を扱う分野の理論研究と応用研究が盛んに行われている。このような計量経済学での研究動向はこの間の経済学における実証研究の興隆というより大きな潮流と表裏一体の関係があると考えられ、近年における計量経済学における一つの重要な流れと理解すべきであろう。

今回は幸いにも、こうしたマイクロ計量経済学の発展に際して、J. Heckman 教授 (シカゴ大学) や D. MacFadden 教授 (カリフォルニア大学バークレー校) と並んで、これまで既に世界的に多くの重要な貢献をされてきているスタンフォード大学経済学科の Takeshi Amemiya (雨宮健) 教授が来日され、東京大学経済学研究科及び CIRJE (日本経済国際共同研究センター) において連続講義を行っていただく機会を得た。おそらくは滅多に無いこの機会をより有意義とするために、ここに Amemiya 教授が行った連続講義の内容を「マイクロ計量経済学」 (Lectures on Microeconometrics) として日本語でまとめることとした。Amemiya 教授によればこれまで研究内容を日本語で説明したものは皆無ということであるが、それにも関わらず今回の講義を日本語で行って頂けたことは特に院生諸君には幸いであつたように思われる。講義の内容はこれまで日本語では説明されたことがほとんどないと思われる高度ではあるが思慮に富んだ内容が多く含まれ、近年に発展がめざましい計量経済学 (Econometrics) やマイクロ計量経済学に関心のある多くの諸氏にとって非常に有用な内容であると考えられる。

なお、この講義録をまとめる作業に当たっては COE プロジェクト「市場機構と非市場機構の関連研究拠点」の援助を受けたことに言及しておく。講義録の作成には講義を聴講した東京大学大学院経済学研究科・統計コースに属する院生 (赤司健太郎・宮脇幸治・松井宗也・松下幸敏・南慎太郎の諸君) 及び国友直人がまとめたメモを元としているが、この講義録の内容については Amemiya 教授の校閲をも経たものである。また、連続講義に先立って国友が Amemiya (1985) の第 3 章と第 4 章にもとづいて Amemiya 教授の講義履修の為に前提となる確率論・数理統計学に関する若干の予備知識の説明を行ったので、その概要を付録としてまとめておいた。

ここで講義録の作成に協力してくれた赤司君をはじめとする院生に対して特に感謝する。

CIRJE・センター長
国友直人

2005 年 6 月

目次

- 第1回：枝分かれロジット (Nested Logit) モデル
- 第2回：トービット (Tobit) モデル
- 第3回：一般化トービット (Generalized Tobit) モデル
- 第4回：デュレーション (Duration) モデル
- 第5回：左打ち切り (Left Censoring) 問題
- 付論：非線型計量分析の為の統計的漸近理論

第1回：枝分かれロジット (Nested Logit) モデル

2005年4月18日

概要

枝分かれロジット・モデル (Nested Logit Model、略して NLM) を応用した実証研究 Amemiya=Shimono (1989)¹ では、高齢者の職業選択 (フルタイム・パートタイム・自営業・退職) の要因を分析している。この例を用いて多項ロジット (Multinomial Logit Model、略して MLM) の定式化や MLM の仮定に含まれる「無関係な選択肢からの独立性」(IIA) を説明する。また、MLM の諸条件 (効用最大化・IIA・二重指数分布の仮定) に関する諸定理、並びに NLM 推定法など、MLM と NLM について理論と応用面から解説する。

1 質的変数モデル (qualitative response model) の定式化

質的変数モデル (qualitative response model) または、離散選択モデル (discrete choice model) は、もともとは生物学・工学・医学等の分野で用いられていた統計的モデルである。例えば、病気に対してある治療法 (treatment) に効果があるか否かは、離散的 (discrete) または質的な (qualitative) な反応 (response) ととらえることができるので、線形回帰モデルにおける従属変数のように数量的な (quantitative) ではない量が観察されると見なすことができる。全体の標本数 n の中である患者 i にとって離散的な反応がある (効果がある) ことを $y_i = 1$ と表し、反応がない (効果がなかった) ことを $y_i = 0$ と表しておく。ここで反応の確率を一般に、

$$P(y_i = 1) = G(x_i^*, \beta) = F(h(x_i^*, \beta)) \quad (1)$$

と表そう。ここで、 G はある累積分布関数、 x_i^* は独立変数、 β は推定したい未知のパラメーター・ベクトルである。先の例では独立変数としては新薬の投入量などが考えられるので量的変数でも構わない。さらに、関数 G を特定化することで反応の確率を定めることができるが、累積分布関数として特定の F を用いることが便利なので、 $h = F^{-1} \circ G$ という変換を行うことによりここで分析に一般性を

¹Amemiya Takeshi and Shimono Keiko (1989), "Application of Nested Logit Models to the Labor Supply of the Elderly," *The Economic Studies Quarterly*, Vol. 40, 14-22.

失うことはあまりないことが分かる。また、変数 x_i^* の確率への影響を母数を利用して表現する場合には $h(\cdot)$ を線形関数 $x_i'\beta$ (説明変数の観測値を x_i とする) を用いることが便利であり、その結果としては

$$P(y_i = 1) = F(x_i'\beta) \quad (2)$$

が得られる。ここで、 F としてはしばしば $\Phi(x)$ (標準正規分布) や $\Lambda(x) = \frac{e^x}{1+e^x}$ (ロジスティック分布) が用いられる。前者をプロビットモデル、後者をロジットモデルと呼ぶ。ここで、モデルの適合度 (fit) を良くするために色々な変換 h を用いることは、質的変量モデル (qualitative response model) に限った特殊なことではないことに注意しておく。例えば線形回帰モデルなどにおいても独立変数の2乗項や交差項を加えたり、対数変換項を用いることはしばしば行われている。概ね、 x を x^* の適当な変換と考えればよい。

以上の例では、選択肢が2つ ($j = 0, 1$) の2値モデル (binary univariate model) であり QRM の範疇では最も簡単なものといえる。次に選択肢 j が3つ ($j = 1, 2, 3$) または、3つ以上の場合に効用最大化² (utility maximization) に基づく定式化を考察してみよう。そこで通勤手段にバス・自家用車・電車の3つの選択の内、いずれを用いれば効用が最大化されるのか、という設定を考える。ある個人 i ($i = 1, \dots, n$) が通勤手段 j ($j = 1, 2, 3$) を用いたときの効用を $U_{j,i}$ とし、確定的な項 $\mu_{j,i}$ と確率的な項 $\epsilon_{j,i}$ の和で表されるものとする。

$$U_{j,i} = \mu_{j,i} + \epsilon_{j,i}, (j = 1, 2, 3) \quad (3)$$

ここで $U_{j,i}$ は観測されないが、効用最大化の条件よりバス ($j = 1$) が選択された場合には $U_{1,i} > \max(U_{2,i}, U_{3,i})$ が成立しているはずである。つまり、ある j が選択される確率は、 $\{U_{j,i} > \max_{k \neq j} U_{k,i}\}$ が起きる確率を評価することで与えられる。

ところで確定的な項 $\mu_{j,i}$ の選択肢 j への依存度により、様々な可能性が生じる。一般には、[1] $\mu_{j,i} = x'_{j,i}\beta_j$ とすることができる。(ただし、 $\beta_{j,i}$ と i にまで依存させるとパラメータ数が多くなりすぎる)。[2] 次に $\mu_{j,i} = x'_{j,i}\beta$ とすると、[1] に比べてモデルの適合度 (fit) は劣るがより効率的な (efficient) 推計が可能となる。[3] $\mu_{j,i} = x'_i\beta_j$ とする定式化は、利用可能なデータ上の制約がある場合に利用されることがある。ここで通勤手段を例に独立変数 $x_{i,j}$ の属性をあげると、個人 i についての属性としては、自宅から勤務先までの距離、交通費、あるいは所得などが考えられる。一方、通勤手段 j への依存は交通費であるが (電車とバスでは公共運賃は異なる)、 j に依存したデータが常に得られるとは限らない。Amemiya=Shimono (1989) では利用可能なデータの制約から [3] による定式化を用いている。

²ただし、例えば殺虫剤が効くか否かというなど設定では、害虫が効用最大化をしているわけではないので、(3) 式は一般に抵抗力 (resistance) や傾向 (tendency) を表現していると考えられるのが適当だろう。こうした解釈は消費者が財を購入する時にも適切な場合がある。

2 多項ロジット・モデル

多項ロジット・モデル (multinomial logit model with 3-responses) を Amemiya=Shimono 論文を例にとって説明すると、職業選択 (フルタイム・パートタイム・自営業・退職) としては4つある。ただし、退職という選択は考慮せず (フルタイム・パートタイム・自営業) の3つの選択肢があるとしよう。ここで [3] による定式化 $\mu_{j,i} = \mathbf{x}'_i \beta_j$ のもとで、各 $j (j = 1, 2, 3)$ が選択される確率は2値モデルの (1) 式と同様に、

$$P(y_i = 1) = G_1(x_i^*, \beta_1) = F_1(h_1(x_i^*, \beta_1), h_2(x_i^*, \beta_2)) \quad (4)$$

$$P(y_i = 2) = G_2(x_i^*, \beta_2) = F_2(h_1(x_i^*, \beta_1), h_2(x_i^*, \beta_2)) \quad (5)$$

と表現される。ここで (4) 式はフルタイム ($j = 1$) を選択する確率、(5) 式はパートタイム ($j = 2$) を選択する確率を表している。 $P(y_i = 3) = F_3(\cdot)$ については定める必要はなく、 $\sum_{j=1}^3 P(y_i = j) = 1$ から $P(y_i = 3) = 1 - F_1(\cdot) - F_2(\cdot)$ により定められるので特定化すべきは $F_1(\cdot)$ 及び $F_2(\cdot)$ となる (つまり j 個の選択肢のうち適当なひとつを除いて $j - 1$ 個の F を特定化すればよい)。2項モデルに関する議論と同様にすると関数 h_1 と h_2 を選ぶことができるので、関数 F_1 と F_2 はできるだけ扱いやすい簡単な形がよいと考えられよう。例えば3反応項目の多項ロジット・モデル (multinomial logit model with 3-response) は、

$$F_1 = \frac{e^{h_1}}{1 + e^{h_1} + e^{h_2}}, F_2 = \frac{e^{h_2}}{1 + e^{h_1} + e^{h_2}}, F_3 = \frac{1}{1 + e^{h_1} + e^{h_2}} \quad (6)$$

により定められる。

これらの (4) 式と (5) 式により各々の選択確率が定まるが、(3) 式の効用最大化問題を考えると $(\epsilon_{1,i}, \epsilon_{2,i}, \epsilon_{3,i})$ が、どのような確率法則に従えば (6) 式または、

$$F_j = P(\{U_j > \max_{k \neq j} U_k\})$$

が言えるのだろうか？この疑問への解答は

「 $\epsilon_{j,i}$ が j と i について独立に二重指数分布 $F(\epsilon) = \exp(-\exp(-\epsilon))$ に従うとき」

ということになる (後述の定理1を参照)。

ここで二重指数分布 (Type I の極値分布 (extreme value distribution)) の特性について簡単に述べておこう。この分布の期待値は $E[\epsilon] \simeq 0.577$ (オイラー定数)、分散 $Var[\epsilon] \simeq 1.645$ である。確率密度関数は

$$f(\epsilon) = \exp(-\epsilon - \exp(-\epsilon))$$

で与えられるが、図1に示したように右に歪んだ (skewed) 形をしている。ところで確率的な項 $\epsilon_{j,i}$ の期待値がゼロではないが、選択確率の推計には影響を与えないことに注意する。線形回帰モデルと同様に、期待値がゼロでなくても定数なら定数項に吸収すればモデルとしては同じことである。

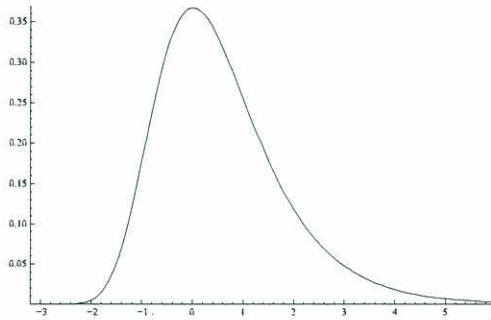


図 1: 二重指数分布の密度関数 : $f(0) \simeq 0.37, f(-2) \simeq f(5) \simeq 0$

多項ロジット・モデル (multinomial logit model) の性質について 2 つ挙げておこう。1 つはロジット・モデル (binary logit) との関連性を示すものであり ($\mu_{j,i}$ の i を省略して μ_j として)、

$$P(y_i = 1 | y_i = 1 \text{ or } y_i = 2) = \frac{e^{\mu_1}}{e^{\mu_1} + e^{\mu_2}} \quad (7)$$

が成り立つことである。これは選択肢が 2 つ (フルタイムかパートタイム) に絞られたとき、条件付の選択確率はロジット・モデルに帰着されることを意味している。

もう 1 つの重要な特性は、「無関係な選択肢からの独立性」 (independence from irrelevant alternatives : IIA と略記) と云われている性質で、これは次の等式

$$P(U_{3,i} > U_{1,i}) = P(U_{3,i} > U_{1,i} | U_{3,i} > U_{2,i} \text{ or } U_{1,i} > U_{2,i}) \quad (8)$$

を意味する。ここで比較している選択肢は、例を用いると自営業 ($j = 3$) とフル・タイム ($j = 1$) の選択に際して、パート・タイム ($j = 2$) は効用 $U_{2,i}$ が低く選ばれることはなかった、ということが条件になっている。IIA は、この条件とは独立に自営業かフル・タイムかの選択確率が定まるという性質である。ここで、給料が少ないからパート・タイムが選ばれなかったとすると、条件からは選択は独立とも解釈されよう。しかしながら、例えば上司に付くのが嫌いな人の雇用の選択においては等式が成り立つことは疑わしい。また、この問題については McFadden による「青いバス (blue bus) と赤いバス (red bus)」の例が有名である。交通手段として 1. 青いバス、2. 赤いバス、3. 自家用車、という 3 つの選択肢があると、赤いバスを利用しない人は青いバスを利用しない傾向にあると考えられよう。この例はかなり人工的であるが経済分析における多項ロジット・モデルや IIA を利用する問題点を鋭く指摘しており、McFadden が枝分かれロジット・モデル (nested logit model) を導入したきっかけでもあった。

そこで IIA を含む多項ロジット・モデル (multinomial logit model) に関する諸条件を整理し、もう少し理解を深めておくことは、枝分かれロジット・モデル

(nested logit model) を理解する上でも有用と考えられる。これまでに現れた5つの諸条件を次のように略記する。

1. UM : 効用最大化 (utility maxmaization),
2. IID : $\epsilon_{j,i}$ (あるいは i を省略し ϵ_j) が同一分布に独立に従う (independent and identically distributed、略して i.i.d.),
3. IIA : Independence from Irrelevant Alternatives, ³
4. MLM : (6) 式 (Multinomial Logit model),
5. EVD : $F(\epsilon) = \exp(-\exp(-\epsilon))$ (極値分布 extreme value distribution)。

このとき次のような結果が成り立つ。

定理 1

UM, IID, EVD \Rightarrow MLM

証明 : Amemiya (1985) ⁴ p.297 の (4.3.42) を参照。

定理 2

UM, IID, EVD \Rightarrow IIA

証明 : 同時分布 (U_1, U_2, U_3) が定まるとき積分を具体的に計算する。

$$\begin{aligned} P(U_1 > U_2) &= \int_{-\infty}^{\infty} F(\mu_1 - \mu_2 + \epsilon) f(\epsilon) d\epsilon \\ &= \int_{-\infty}^{\infty} \exp(-\exp(-(\mu_1 - \mu_2 + \epsilon))) \exp(-\epsilon) \exp(-\exp(-\epsilon)) d\epsilon \end{aligned}$$

より、 $\nu = \epsilon + \mu_1$ と変換すれば

$$\begin{aligned} &= \int_{-\infty}^{\infty} \exp(-\exp(-(-\mu_2 + \nu))) \exp(-(\nu - \mu_1)) \exp(-\exp(-(\nu - \mu_1))) d\nu \\ &= \frac{e^{\mu_1}}{e^{\mu_1} + e^{\mu_2}} \int_{-\infty}^{\infty} e^{-\nu} (e^{\mu_1} + e^{\mu_2}) \exp(-e^{-\nu} (e^{\mu_1} + e^{\mu_2})) d\nu \\ &= \frac{e^{\mu_1}}{e^{\mu_1} + e^{\mu_2}} \end{aligned}$$

となる。他方、条件付の選択確率は3つの事象 $A \equiv \{U_1 > U_2\}$, $B \equiv \{U_1 > U_3\}$, $C \equiv \{U_2 > U_3\}$ と置くと、

$$P(U_1 > U_2 | U_1 > U_3 \text{ or } U_2 > U_3) = P(A \text{ and } (B \text{ or } C)) / P(B \text{ or } C)$$

であるが、分母は

$$P(B \text{ or } C) = 1 - P(\bar{B} \text{ and } \bar{C}) = 1 - P(y_j = 3) = \frac{e^{\mu_1} + e^{\mu_2}}{1 + e^{\mu_1} + e^{\mu_2}}$$

³3つの選択肢がある場合、 i を略して U_j ($j = 1, 2, 3$) に対して $P(U_1 > U_2) = P(U_1 > U_2 | U_1 > U_3 \text{ or } U_2 > U_3)$, $P(U_1 > U_3) = P(U_1 > U_3 | U_1 > U_2 \text{ or } U_3 > U_2)$, $P(U_2 > U_3) = P(U_2 > U_3 | U_2 > U_1 \text{ or } U_3 > U_1)$ の3式から成ることに注意。

⁴Amemiya, Takeshi (1985), *Advanced Econometrics*, Blackwell.

分子は、

$$\begin{aligned} P(A \text{ and } (B \text{ or } C)) &= P(A \text{ and } B) + P(A \text{ and } C) - P(A \text{ and } B \text{ and } C) \\ &= P(A \text{ and } B) = P(y_j = 1) = \frac{e^{\mu_1}}{1 + e^{\mu_1} + e^{\mu_2}} \end{aligned}$$

となる。したがって、(分子)/(分母)により結果が示される。

定理 3

UM, IID, IIA \Rightarrow EVD

[注意]：この定理3はこれまでの文献上には見あたらないようである。証明は以下の配布資料を参照。

定理 4

UM, IID, MLM \Rightarrow EVD

証明：McFadden (1974)⁵を参照。

以上で述べた諸定理には二つの対応関係がある。定理3はUM, IIDの条件の下で、定理2の逆 (inverse) と捉えることができる。同様に定理4は定理1の逆に対応している。

ここで配布資料に基づき定理3の証明の概要を述べておく。

[定理3：UM, IID, IIA \Rightarrow EVD の証明]

X, Y, Z が独立に分布 $F(x), G(y), H(z)$ に従っているとすると、

$$\begin{aligned} P(Y > X) &= \int_{-\infty}^{\infty} F(y)g(y)dy, \\ P(Y > X | Z > Y \text{ and } Z > X) &= \frac{\int_{-\infty}^{\infty} [\int_{-\infty}^z F(y)g(y)dy]h(z)dz}{\int_{-\infty}^{\infty} F(z)G(z)h(z)dz} \end{aligned} \quad (9)$$

が得られる。

[注意点：(9)式では、 $P(Y > X) = P(Y > X | Z > Y \text{ and } Z > X) \Leftrightarrow P(Y > X) = P(Y > X | Y > Z \text{ or } X > Z)$ を用いている。 $\{Y > X\}$ は補集合 $\overline{\{Y > Z \text{ or } X > Z\}} = \{Z > Y \text{ and } Z > X\}$ とも独立となる。]

したがって、 $P(Y > X) = P(Y > X | Y > Z \text{ or } X > Z)$ は

$$\int_{-\infty}^{\infty} [\int_{-\infty}^z F(y)g(y)dy - \int_{-\infty}^{\infty} F(y)g(y)dy \cdot F(z)G(z)]h(z)dz = 0$$

⁵McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics* P. Zarembka ed., 105-142, Academic Press.

で与えられる。ここで、 $G(y) = F(y + a)$, $H(z) = F(z + b)$ を仮定すると (IIDは ϵ_j の分布についてであり U_j は同一分布には従わない。しかし (2) 式より分布の位置 (location) のみが異なっている) 上式は

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^z F(y)f(y+a)dy - \int_{-\infty}^{\infty} F(y)f(y+a)dy \cdot F(z)F(z+a) \right] h(z)dz = 0$$

となる。

したがって、大括弧 $[\cdot]$ の中は、任意の z に対してゼロでなければならない。 $\int_{-\infty}^z F(y)f(y+a)dy = cF(z)F(z+a)$ (ただし定数 $c = \int_{-\infty}^{\infty} F(y)f(y+a)dy$) とおいて、 z について両辺を微分すると、

$$F(z)f(z+a) = c[f(z)F(z+a) + F(z)f(z+a)]$$

が得られる。さらに両辺を $F(z)$ で割り整理して、

$$\frac{f(z)}{F(z)} = c_1 \frac{f(z+a)}{F(z+a)}$$

となる。

すなわち、 $\frac{\partial}{\partial z} \log F(z) = \frac{\partial}{\partial z} c_1 \log F(z+a)$ (ここで $c_1 = \frac{1}{c} - 1$ である)。同じく、任意の z に対して $\log F(z) = c_1 \log F(z+a)$ または、 $-\log F(z) = c_1 [-\log F(z+a)]$ と表せる。さらに対数をとって、 $\log[-\log F(z)] = \log c_1 + \log[-\log F(z+a)]$ なので z について微分することで、

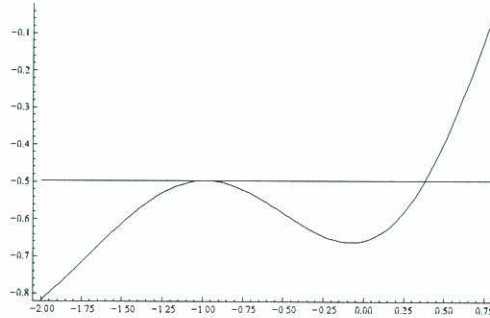


図 2: 導関数 $h'(z)$ について。例えば、 $h'(z_0 = -1) \neq h'(z_0 + a = 0.75)$

$$\frac{\partial}{\partial z} \{ \log[-\log F(z)] - \log[-\log F(z+a)] \} = 0 \quad (10)$$

より (10) 式を得る。ここで、 $h(z) = \log[-\log F(z)]$ とおくと、任意の z に対して (10) 式が成り立つためには、 $h(z)$ は z に関して線形 (linear) でなければならない (図 2 参照)。つまり、

$$\log[-\log F(z)] = cz + d$$

したがって、 $-\log F(z) = e^{cz+d}$ より、二重指数分布関数

$$F(z) = \exp(-\exp(cz + d)) \quad , (c < 0) \quad (11)$$

が得られる。

3 枝分かれロジット・モデル

多項ロジット (multinomial logit model) では、先見的に IIA が成立していることがわかった。そこで、MLMを拡張して、IIA の仮定を克服しようとするモデルが枝分かれロジット・モデル (nested logit model、略して NLM) である。

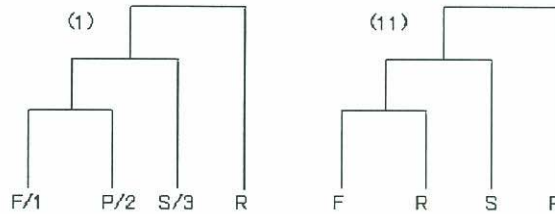


図 3: 入れ子構造 (Amemiya=Shimono 論文):フルタイム (F と表示)、パートタイム (P)、自営業 (S)、退職 (R)

図 3(の左図) は、前節と同じく選択肢が 3 つ ($j = 1, 2, 3$) の場合についての入れ子 (nest) 構造を示している。このとき、確率的な項 ϵ_j の同時分布は $F(\epsilon_1, \epsilon_2) \cdot F(\epsilon_3)$ となる。 $F(\epsilon_3)$ は MLM でも用いられた二重指数分布 (Type 1 の極値分布 extreme value distribution) であるが、

$$F(\epsilon_1, \epsilon_2) = \exp(-[e^{-\frac{\epsilon_1}{\rho}} + e^{-\frac{\epsilon_2}{\rho}}]^\rho) \quad , 0 < \rho \leq 1 \quad (12)$$

は Gumbel's Type-B のグンベルの 2 次元極値分布 (bivariate extreme value distribution) と呼ばれている。ここで、パラメーター ρ は相関係数と関係しており、 $Corr(\epsilon_1, \epsilon_2) \simeq 1 - \rho^2$ となることが知られている。また、 $\rho = 1 \Leftrightarrow$ 独立となり ((12) 式で $\rho = 1$ とすると $F(\epsilon_1) \cdot F(\epsilon_2) \cdot F(\epsilon_3)$)、 $\rho = 1$ のときは MLM に還元される。 $(\rho > 1$ のときには密度関数は負値をとることがある。) 従って、図 3(の左図) の NLM は、選択肢 (1,2) と選択肢 3 は相関のない状態、1 と 2 は相関がある状態を想定していることになる。

この問題における尤度関数を表現するため、3 つの選択肢についての観測値を $y_i = j$ ($j = 1, 2, 3$) $\Rightarrow y_{j,i} = 1$ (その他のときはゼロ) と置くことにしよう。尤度関数 L の構成は MNL と同様に、

$$L = \prod_i P(y_{j,i} = 1)^{y_{1,i}} P(y_{j,i} = 2)^{y_{2,i}} P(y_{j,i} = 3)^{y_{3,i}}$$

表 1: 配布資料の抜粋 (Amemiya=Shimono 論文・Table 1)

		$-\log L$	Estimate of ρ
Multinomial Logit		2659.7	—
Nested Logit Type I	(1)	2650.1	6.41(6.49)
	(2)	2653.1	4.96(8.44)

	(11)	2660.8	.98(.37)
	(12)	2658.0	.50(.24)

括弧内は標準誤差。

$$\begin{aligned}
 &= \prod_i P(y_{j,i} = 1 | y_{j,i} = 1 \text{ or } y_{j,i} = 2)^{y_{1,i}} P(y_{j,i} = 2 | y_{j,i} = 1 \text{ or } y_{j,i} = 2)^{y_{2,i}} \quad (13) \\
 &\quad \times \prod_i P(y_{j,i} = 1 \text{ or } y_{j,i} = 2)^{y_{1,i} + y_{2,i}} P(y_{j,i} = 3)^{y_{3,i}} \\
 &= L_1 \times L_2
 \end{aligned}$$

と表せる。ここで尤度関数は条件付確率により $L_1 \times L_2$ に分解できることに注目する。現在の最大化アルゴリズムでは尤度関数 L を直接に最大化することができるが、Amemiya=Shimono 論文の当時は推計が困難であった。そこで、パラメーターを 2 段階に分けて推計する方法として

$$\text{第 1 段階 : } L_1 \text{ を } \frac{\beta_1 - \beta_2}{\rho} \text{ に関して最大化} \quad (14)$$

$$\text{第 2 段階 : } L_2 \text{ を } (\beta_2 - \beta_3) \text{ と } \rho \text{ に関して最大化} \quad (15)$$

が考えられる。ここでの第 1 段階は Logit プログラムによる推定で、第 2 段階で最終的に識別したいパラメーターの推計を行っている。しかしながら、二段階に分けることで推定の漸近的な効率性は落ちるので、直接の最尤法に拠ったほうがより望ましいと言えよう。

さて Amemiya=Shimono 論文の実証例では、先に述べた通り高齢者の職業選択は 4 つと想定されている。MLM と NLM による推計結果の抜粋⁶を表 1 に示してある。表 1 の Type I(1) と Type I(11) は、先の図 3 の入れ子構造 (1) と (11) に対応しており、 $-\log L$ の値が最も小さいか、大きい場合に当たる。(1) の場合、フルタイムとパートタイムは近い状態と想定され、 $-\log L$ の値が小さく直感にかなう。一方、フルタイムと退職をまとめた (11) では $-\log L$ の値が大きい。

MLM と NLM の比較では、パラメーター ρ に着目する必要がある。表 1 の Type I(1) で、 ρ の推定値は 1 より大きく離れているが、標準誤差も大きくなっている。推定された尤度などから Amemiya=Shimono 論文では、MLM が許容されるという実証結果となっている。さらに、最尤法による再推計を行ったところ、 $\rho = 1$

⁶Amemiya and Shimono (1989) では MLM と NLM に関し 4 タイプで合計 25 通りの質的変量モデルを推定している。

とは有意に異なるという結果が得られた。しかし、推定値が $\hat{\rho} > 1$ であったため、 $(0 < \rho \leq 1$ のときに)MU(効用最大化)から選択確率が導かれる、という条件とは整合的ではなかった。これまで報告されている多くの実証結果では同様の推定結果が報告されているが、それらの経済学的解釈は簡単ではない。こうした問題を解決するためにこれまで幾つかの研究がなされているが、例えば Amemiya and Kim (2002)⁷ を見よ。一つの可能な解釈は、これを純粋な統計モデルとしてみなすことである。 $\rho > 1$ でも確率は常に 0 と 1 の間に入り、その和は 1 になるので統計学的矛盾は生じない。

⁷Amemiya and Kim (2002), "A Generalization of the Nested Logit Model," in Klein and Mitlnik, eds, *Contributions to Modern Econometrics*, Kluwer Academic Publishers.

第2回：トービット (Tobit) モデル

2005年5月11日

概要

Tobin(1958)¹ は耐久財の消費に関する実証分析に際して新たな計量モデルを提案した。後にトービット・モデルと呼ばれるこのモデルでは、従属変数 (dependent variable) の観測値の一部が打ち切られる (censored) 特長を有する。この種の計量経済モデルの定式化や推計法 (最尤法/Heckman 法/EM アルゴリズム) を解説する。また、誤差項の仮定を緩めた場合 (非正規性・不均一分散・系列相関) の推定量の性質を概観した後、Kaplan-Meier 推定量を用いたセミパラメトリック推計法も紹介する。

1 トービット・モデルの定式化と推計法

実数値をとる被説明変数の計測データがある閾値を境に観測されなくなる場合に関しては、特に元の潜在変数が i.i.d. (independent and identically distributed) にしたがっていると想定した分析が生物統計の分野で様々な研究がなされていた。回帰分析に対してこうした状況を適用し、計量経済分析をはじめて行ったのは Tobin(1958) の耐久財 (durable goods) に関する消費の分析においてである。ここでまず、Tobin(1958) の元々の分析を例にトービット・モデルの定式化と推計法について解説する²。

¹Tobin, James (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, Vol. 26, 24-36.

²ここで米国の計量経済学の動向を簡単に回顧してみよう。1970年頃まではマクロ計量モデルなどを用いた経済予測が主流であった。しかしながら、計量経済モデルを用いた経済予測の結果があまり思わしくないことが議会で問題となったことや、1970年頃から各地で大々的にクロスセクション・サーベイが盛んに行われるようになったことなどを契機に、時系列分析からクロスセクション分析の流れが顕著となっていった。こうしたことが今回と次回にトービット・モデル等を用いる計量分析の重要性の背景である。

ある消費者 i の耐久財に対する所期支出額 (desired expenditure) を y_i^* と表すことにする。 y_i^* は、耐久財の需要関数を x_i^* 、耐久財価格 p とすれば、 $y_i^* = p \cdot x_i^*(I_i)$ と表され所得 I_i (Income) の関数となる。単回帰モデルを考え、独立変数を所得 ($x_i \equiv I_i$) とすれば、

$$y_i^* = \alpha + \beta x_i + u_i \quad (1)$$

となる。ここで、 α, β は経済学的に意味のある推定したい母数、 u_i は誤差項である。ここで問題となるのは、所期支出 y_i^* が負の値の時³ には、耐久財の購入量は正に限られるので、観測される y_i^* の値はゼロとなることである。 y_i^* が正のときには観察され、負の値のときには観察されずに、ある期間において購入量がゼロとなっている (y_i^* もゼロ) 状況を図1に示しておこう。

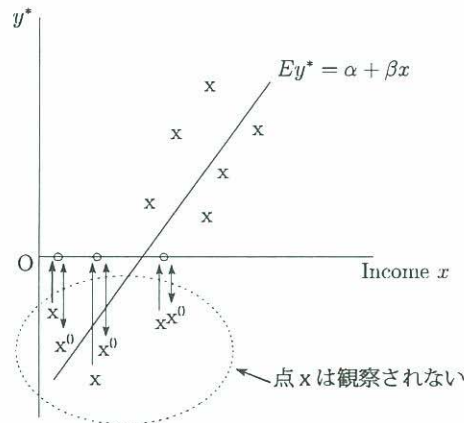


図 1: 打ち切りデータを伴う回帰分析 (点 x^0 については後述)

実際に観察される y_i^* の値を y_i とすると、観測データがしたがうルールは

$$y_i = y_i^* \text{ if } y_i^* > 0 \quad (2)$$

$$y_i = 0 \text{ if } y_i^* \leq 0 \quad (3)$$

となる。(この例では閾値をゼロとしたが任意の水準でよい。) この観測ルールの下で、(1) 式の単回帰式を最小二乗法 (OLS) により推計するなら

³潜在変数 y_i^* が所期支出となる例においては「 y_i^* 自体がそもそも負の値をとり得るのか」という問が考えられる。ここでは、「もしも1万円をくれるなら冷蔵庫をもらう」、「1万円なら耐久財を売りたい」と解釈することにしよう。

ば、次のようなことが言える。まず、すべての y_i を用いて (つまり、 y_i^* が負で観測されないデータは $y_i = 0$ となっている) 回帰すると、図1から真の傾き β よりも、緩やかな傾き $\bar{\beta}$ をもつ回帰直線を当てはめてしまうことが想像される。他方、正の値をとる $y_i > 0$ のみを用いて回帰を試みればゼロのデータを無視するのでこれも偏り (bias) を生じる。(後述の (11) 式を参照。)

1.1 トービット尤度関数 (Tobit LF) の導出

トービット・モデルの最小二乗推定による偏りは、(2),(3) 式の観測ルールを考慮していないことに起因している。すなわち、(2),(3) 式から $y_i = 0 \Rightarrow y_i^* \leq 0$ である (事象 $\{y_i^* \leq 0\}$ は観測できる)。そこでこれを y_i^* が既知となる区間 $(-\infty, 0] = S_i \subset \mathbb{R}$ に入っていること ($y_i^* \in S_i$) と理解し、 y_i^* の密度を f_i と表しておき、以下のように尤度を一般化する。

$$L = \prod_{i=1}^n \int_{S_i} f(y_i^*) dy_i^* \quad (4)$$

この尤度では、約束事としても S_i が一点集合 (つまり、 $S_i = \{y_i\}$ と $y_i^* = y_i$ が正の値で観測された) のときは、密度 $f(y_i) \equiv \int_{S_i} f(y_i^*) dy_i^*$ で評価することとする。このとき、例えば常に $y_i^* > 0$ とすれば、尤度は $L = \prod_{i=1}^n f(y_i)$ となり通常の尤度関数の形となる。

トービット・モデルの尤度関数 (Tobit Likelihood Function: Tobit LF と略記) を (4) 式に従って構成すると、

$$\begin{aligned} L &= \prod_{i|y_i^* > 0} f_i(y_i) \prod_{i|y_i^* \leq 0} P(y_i^* \leq 0) \\ &= \underbrace{\prod_0 P(y_i^* \leq 0) \prod_1 P(y_i^* > 0)}_{\text{LF of Probit}} \times \underbrace{\prod_1 f_i(y_i | y_i^* > 0)}_{\text{LF of Truncated regression model}} \end{aligned} \quad (5)$$

で与えられる。上式のように密度と確率からなる尤度は、一般化された尤度関数 (Generalized likelihood function) と呼ばれ、 S_i を定めることで、どのような場合でも尤度が組めることを示唆している。一般化された尤度

に基づく Tobit MLE (Tobit LF に基づく最尤推定量) の漸近的性質と正規条件 (regularity condition) については、Amemiya(1973)⁴が導いている。

ここで Tobit LF の構成をみると 1 番目の要素は、プロビット・モデル⁵と切断回帰 (truncated regression) モデルの尤度の積で表わされていることがわかる。トービット・モデルの推計法に関係するが、 L (プロビットの LF を含む) に拠る推計は「打ち切り回帰モデル (Censored regression model)」に属し、切断部分の尤度のみを拠る推計は「切断回帰モデル (Truncated regression model)」と呼ばれ、区別する必要がある。両者の違いは、後者においては $y_i^* < 0$ の時に、独立変数 (外生変数) x_i も観察されるか否かという違いにある。

尤度関数を構成する第二の項、Tobit LF における密度関数 f の特定化についてであるが、Tobin(1958) では誤差項が正規分布に従うと仮定している。これを標準的なトービット・モデル (standard Tobit model) と呼ぶが、次のようにまとめることができる。

$$y_i = \max\{y_i^*, 0\}, \quad i = 1, 2, \dots, n \quad (6)$$

$$y_i^* = \mathbf{x}_i' \beta + u_i, \quad u_i \sim N(0, \sigma^2) \quad (7)$$

1.2 Heckman の 2 段階推計

トービット・モデルの推計法としては Heckman(1976)⁶ による 2 段階の推計法がよく知られている。この方法の第 1 段階は、(5) 式の LF of Probit に拠る推計で、第 2 段階は、同じく (5) 式の Truncated regression model に基づく手順から構成されている。すなわち第 1 段階では、

$$\text{Probit ML} : \prod_0 [1 - \Phi(\mathbf{x}_i' \beta / \sigma)] \prod_1 \Phi(\mathbf{x}_i' \beta / \sigma) \quad (8)$$

となる⁷ プロビット・モデルから母数 $\alpha \equiv \beta / \sigma$ の最尤推定値 $\hat{\alpha}$ を得ることが出来る。ただし、ここでの注意点としてはこの段階では β と σ は区

⁴Amemiya (1973), "Regression Analysis when the Dependent Variable is Truncated Normal," *Econometrica*, Vol. 41(6), 997-1016. において、後述の standard Tobit MLE の一致性と漸近正規性が示されている。

⁵Tobit の名の由来はこの Probit 項との関連から、A. Goldberger により名づけられた。

⁶Heckman (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, Vol. 5, 475-492.

⁷ $u_i \sim N(0, \sigma^2)$ より、 $P(y^* > 0) = P(x' \beta + u > 0) = P(u / \sigma > -x' \beta / \sigma) = \Phi(x' \beta / \sigma)$, $P(y^* \leq 0) = 1 - \Phi(x' \beta / \sigma)$ を得る。ここで、 $\Phi(\cdot)$ は標準正規分布関数。

別して推定できないこと、また2値情報 ($0 : y_i^* \leq 0$ 、または $1 : y_i^* > 0$) から α を推計したので、 β と σ の推定の効率性が損なわれることなどが挙げられる。

続いて、第二段階では (5) 式の切断された尤度 (LF) 部分に拠ることで β と σ を分離して推定するために回帰を考える。まず、切断された尤度 (LF) 部分では $y_i^* > 0$ (つまり、 $y_i > 0$) が観察されているので、(7) 式 of 回帰 (すなわち条件付期待値) は、

$$E(y_i | y_i > 0) = x_i' \beta + E(u_i | u_i > -x_i' \beta) \quad (9)$$

となる。

ここで、 $-x_i' \beta$ を想定した定数を $-\mu$ としておき、 $E(u_i | u_i > -\mu) = \int_{-\mu}^{\infty} u f(u) du / P(u > -\mu)$ を具体的に求める。 u は正規分布にしたがっている⁸、 u と $f(u)$ の積の積分は簡単に求めることができ、

$$\begin{aligned} \text{(分子)} &= \int_{-\mu}^{\infty} u f(u) du & (10) \\ &= \int_{-\mu}^{\infty} \frac{u}{\sqrt{2\pi}\sigma} e^{-\frac{u^2}{2\sigma^2}} du \\ &= \left[-\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}} \right]_{-\mu}^{\infty} = \sigma \phi\left(\frac{\mu}{\sigma}\right) \end{aligned}$$

ここで $\phi(\cdot)$ は標準正規密度である。他方、(分母) は $P(u/\sigma > -\mu/\sigma) = \Phi(\mu/\sigma)$ となるので、結局 (9) 式は、

$$E(y_i | y_i > 0) = x_i' \beta + \frac{\sigma \phi(x_i' \beta / \sigma)}{\Phi(x_i' \beta / \sigma)} \quad (11)$$

と表される⁹。ミル比率 (Mills' ratio) の逆数にあたる $\frac{\phi(x_i' \beta / \sigma)}{\Phi(x_i' \beta / \sigma)}$ を、第1段階のプロビット係数 $\alpha (\equiv \beta / \sigma)$ を用いて表すとしよう。すると (11) 式から、

$$\text{(Truncated regression)} : y_i > 0, y_i = x_i' \beta + \frac{\sigma \phi(x_i' \alpha)}{\Phi(x_i' \alpha)} + \epsilon_i \quad (12)$$

を得る。ここで $\epsilon_i \equiv y_i - E(y_i | y_i > 0)$ である。

第2段階では第1段階のプロビットモデルで得られた推定値 $\bar{\alpha}$ を代入して、ミル比率の逆数を1つの独立変数 (regressor) と見なす。そして、正

⁸正規分布を仮定しているので $E(u_i | u_i > -\mu)$ を明示できることとなる。

⁹このことから例においては $y_i > 0$ のみを用いて回帰項 $E(y_i | y_i > 0) \equiv \alpha + \beta x_i$ を推定すると、 x_i とミル比の逆数に相関があることから偏りが生じてしまう。

の値をとった y_i を $(\mathbf{x}_i, \phi(\mathbf{x}'_i\tilde{\alpha})/\Phi(\mathbf{x}'_i\tilde{\alpha}))$ に回帰し、係数 (β, σ) を推定することとなる。Heckman の推計法は、最小二乗法を用いて簡単に (β, σ) の一致推定量を得る方法と見なすことができる。

一方で、すべてのデータ n ($y_i = 0$ を含む) を用いた回帰も考えられよう。この場合、(11) 式に対応して無条件の期待値、

$$E(y_i) = \Phi(\mathbf{x}'_i\beta/\sigma)\mathbf{x}'_i\beta + \sigma\phi(\mathbf{x}'_i\beta/\sigma) \quad (13)$$

が導出され、この回帰においても一致推定量を得ることができる。ただし、漸近共分散の大小(行列の意味で)は真の母数に依存することになるので、部分的なデータ ($y_i > 0$) か、それともすべてのデータを用いた方が望ましいのかは一概には言えない。

上記2つの2段階推定量と Tobit MLE との比較という点では、モンテカルロ実験の結果から、2段階推定量はかなり非効率的 (inefficient) であることが示されており、これよりトービットモデルの設定では、 ϵ_i の3次以上の積率 (moment) に多くの情報が含まれていると推測される。ただし、米国政府がマイクロサーベイデータを整備しはじめた70年代当時、Tobit MLE(より複雑なトービットモデルを含めて) を求めるには、計算機の能力の制約があったことを指摘しておく。

1.3 EM アルゴリズムを用いた推計法

標準的な Tobit MLE $\tilde{\theta} \equiv (\tilde{\beta}, \tilde{\sigma}^2)$ を求める繰り返し計算法 (Iterative method) として、EM アルゴリズムはトービットモデルに適用する上で解りやすいであろう (詳しくは、Amemiya(1985) ¹⁰ p.375 を参照)。初出の図1に立ち返ってみると、潜在変数 y_i^* が負のときに観察されないことが問題であった。そこで、 $y_i = 0$ となるデータに関しては負の領域へ移し直して(予測値を与え)、あたかも観測データ y_i^0 として扱えないかということを考える。

E-ステップ: まず、初期値 $\theta_{r-1}(r=1,2,\dots)$ を与え、 θ_1 について、潜在変数 \mathbf{y}^* の対数尤度に対して条件付き期待値をとる、 $Q(\theta|\theta_1) \equiv E_{\theta_1}[\log L(\mathbf{y}^*|\theta)|\mathbf{y}]$ 。このとき、 y_i^0 は、

$$y_i^0 \equiv E_{\theta_1}(y_i^*|y_i^* < 0) = \mathbf{x}'_i\beta_1 - \frac{\sigma_1\phi(\mathbf{x}'_i\beta_1/\sigma_1)}{1 - \Phi(\mathbf{x}'_i\beta_1/\sigma_1)} \quad (14)$$

¹⁰Amemiya, Takeshi (1985) *Advanced Econometrics*, Blackwell.

となる。図 1 において、 $y_i = 0$ となり打ち切られていたデータは、傾き β_1 の下、回帰直線の負の領域周辺、点 $x^0(y_i^0, x_i)$ に移ったのではないかと期待される。

M-ステップ: $Q(\theta|\theta_1)$ を θ について最大化し、最大化点 θ_2 を得る。ここで β_2 を見てみると、 $\beta_2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{y}' \mathbf{y}^0]'$ (\mathbf{X} は $(n \times k)$ のデータ行列、 \mathbf{y}^0 は y_i^0 からなる) となり、OLSE の形となる。

そして、EM アルゴリズムに関して、 $LF(\theta_{r+1}) \geq LF(\theta_r)$ ($r = 1, 2, \dots$) が知られているので、EM-step を繰り返すことで Tobit 尤度関数が最大化され、Tobit 最尤推定値 (MLE) $\hat{\theta}$ を得ることが出来る。

2 セミパラメトリック・トービットモデル

2.1 標準的な仮定が満たされないとき

上述の標準的なトービットモデル (standard Tobit model) では、誤差項 u_i に正規分布を仮定してきた。この仮定を緩めた場合でも Tobit MLE は、古典的回帰モデルにおける最小二乗推定量 (OLSE) と同じような頑強性を有するであろうか。ここでの頑強性とは、例えば非正規性・不均一分散・系列相関の下でも OLSE の一致性は保持されるということである。ここで Tobit 最尤推定量 (MLE) の一致性については方程式 $\frac{\partial \log L}{\partial \theta} = 0$ の解が真の確率分布の下で

$$E \frac{\partial \log L}{\partial \theta} = 0$$

となるか否かを調べなければならない。前提となる仮定が満たされない場合には多くの場合には上の式は漸近的にも満たされないことが知られているが、そのことを正確に議論するのはしばしば面倒な問題となる。ここでは Heckman の推計法 (12) 式を利用して推定量の頑健性について直観的に考察してみよう。

(i) 非正規性 (Nonnormality): ミル比率の逆数 $f(\mathbf{x}'\beta, \sigma)/(1 - F(\mathbf{x}'\beta, \sigma))$ の関数形が異なるので、一致性をもたないであろう。

(ii) 不均一分散 (Heteroscedasticity): $\sigma_i \phi(\mathbf{x}'_i \beta_i / \sigma_i) / \Phi(\mathbf{x}'_i \beta_i / \sigma_i)$ が σ_i ($i=1, 2, \dots$) に依存するが、プロビットモデルでは $\alpha = \beta / \sigma$ として推定するので、一般には一致性は望めない。ただし、例えば不均一構造をグループ化して 2 段階推計できれば正しく推定できる可能性がある。

(iii) 系列相関¹¹ (Serial Correlation)：系列相関を有する場合については、一致性が保たれることがRobinson(1982)¹²によって示されている。

従って以上から、トービットモデルにおいては、古典的回帰モデルにおける仮定の重要性が窺われる。

2.2 トービットモデルにおける Kaplan=Meier 推定量

誤差項 u_i が正規分布に従わなければ、標準的な Tobit LF は定式化を誤ってしまう。そこで、誤差分布をノンパラメトリックに推定することを考えよう。以下では β を所与としておくと、トービットモデルでは、

$$(1) y_i^* = x_i'\beta + u_i > 0 \Rightarrow u_i = y_i - x_i'\beta \text{ を観察}$$

$$(2) y_i^* = x_i'\beta + u_i \leq 0 \Rightarrow u_i \leq -x_i'\beta \text{ を観察}$$

が成り立つ。(1) のように $y_i (> 0)$ が1つの値に決まるならば、誤差分布のノンパラメトリック最尤推定量 (MLE) として経験分布関数 (Empirical Distribution Function) を用いることができる。例えば $n = 3$ のとき仮に (母数が既知であり) 誤差項が観察できると考えると経験分布関数 $F(u) = \#\{u_j \leq u\}/3$ を図2のように表現しておこう。

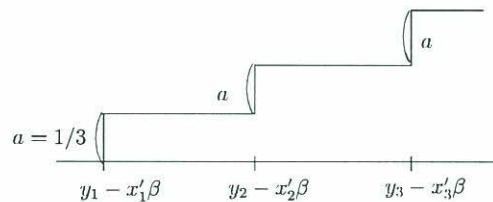


図 2: 経験分布関数

ここで (2) の示す通り打ち切り (censoring) がある場合には、誤差項が区間 $(-\infty, -x_i'\beta]$ に入ることを観察しているので、尤度 (LF) を最大化する分布関数 \hat{F} の構成は明らかでない。そこで $n = 3$ の場合を例に以下のような設定を与えてみよう。

¹¹ 「系列相関」と云うと時系列に限るかのような印象もあるが、クロスセクション間 (州や国家間) では空間的相関を持つモデルも考えることができる。

¹² Robinson, P.M. (1982), "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables," *Econometrica*, Vol. 50, 27-41. は相関構造について (7) 式の u_i が正規定常過程 (stationary gaussian process) にしたがる場合には一致性が保たれることを証明した。

今、仮に $i = 1, 2$ は一義的に (1) が成立し $i = 3$ については (2) が成立するとする。 $y_1 - x'_1\beta < y_2 - x'_2\beta$ の下では、 $\{y_1 - x'_1\beta, y_2 - x'_2\beta, -x'_3\beta\}$ に対する順序の組は 3 通りある。まず図 3 の状況 (打ち切りデータが中間に位置するとき) を仮定しよう。

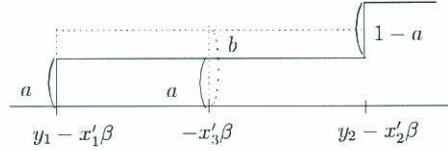


図 3: 順序 1: $y_1 - x'_1\beta < -x'_3\beta < y_2 - x'_2\beta$ (点線は、確率 (高さ) での評価を意味する)

先に結論を述べると、この場合 \hat{F} は必ず図 3 に描かれた通りになる。このとき、 $LF_1(a) = a^2(1-a)$ 、 $\max LF_1 = LF_1(2/3) = 4/27$ ($\partial LF_1/\partial a = 2a - 3a^2 = 0$ より) である。まず、点 $y_1 - x'_1\beta$ では観測点 (uncensored observation) なので増分 (jump) $a = a - 0$ をもち、同時に確率 (高さ) でもある。点 $-x'_3\beta$ は打ち切りデータなので確率で評価し、点 $y_1 - x'_1\beta$ でまた増分をもつ。

ここで重要なことは、打ち切りデータは直前の観測点 $y_1 - x'_1\beta$ での確率で評価されなければ、尤度が最大化されていない、ということである。なぜならば、点 $-x'_3\beta$ を $b (> a)$ で評価してみると $LF_1(a, b) = ab(1-b)$ であるが、そうすると結局、点 $y_1 - x'_1\beta$ の増分を高さ b まで伸ばさなければならなくなる ($\partial LF_1(a, b)/\partial a = b(1-b) > 0$ より)。また、図 2 と異なり¹³、観測点での増分が等しいとは限らないこと ($a \neq 1-a$) がわかる。

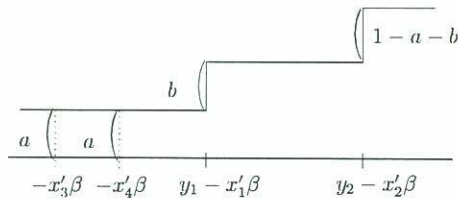


図 4: 順序 2: $-x'_3\beta < y_1 - x'_1\beta < y_2 - x'_2\beta$

次に、図 4 (打ち切りデータ点 $-x'_3\beta$ の順序が最初にくるとき) では、確率 a で評価しているが、仮に 4 つ目の打ち切りデータ点 $-x'_4\beta (< y_1 - x'_1\beta)$ を

¹³通常の経験分布関数では $LF_0(a) = a^2(1-2a)$ より $\max LF_0 = LF_0(1/3)$ で与えられる。

加えても、同様の議論で、等しく確率 a で評価しなければならないことを示している。すなわち $LF_2(a, b) = ab(1 - a - b)$, $\max LF_2 = LF_2(1/3, 1/3) = 1/27$ である。

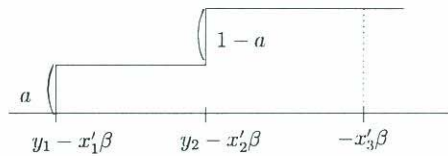


図 5: 順序 3: $y_1 - x_1\beta < y_2 - x_2\beta < -x_3\beta$

最後に図 5(打ち切りデータ点が最も大きいとき)であるが、確率 1 での評価なので、打ち切りデータは尤度に貢献しない。ここで、 $LF_3 = a(1 - a)$ 、 $\max LF_3 = LF_3(1/2) = 1/4$ となる。以上の分析においては β は既知として考えている。 β が未知の場合の推定方法は 2.3 節において論ずる。

以上における尤度の構成法に従うと、一般的な場合は Kaplan=Meier(1958)¹⁴ 推定 (Product limit method) として知られているが、この方法は以下の手続きで行う。

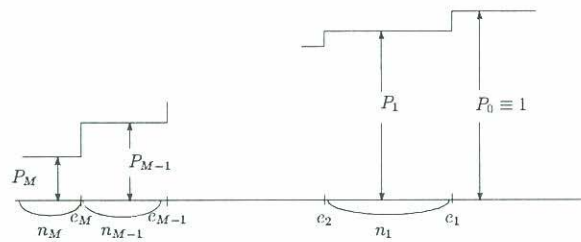


図 6: Kaplan=Meier 推定量 \hat{F}

¹⁴Kaplan=Meier (1958), "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, Vol. 53, 457-481.

$$\begin{aligned}
LF &= P_M^{n_M}(P_{M-1} - P_M)P_{M-1}^{n_{M-1}}(P_{M-2} - P_{M-1})\cdots \\
&\quad \times P_3^{n_3}(P_2 - P_3)P_2^{n_2}(P_1 - P_2)P_1^{n_1}(P_0 - P_1) \\
&= \left[\frac{P_M}{P_{M-1}}\right]^{n_M} \left[1 - \frac{P_M}{P_{M-1}}\right] \left[\frac{P_{M-1}}{P_{M-2}}\right]^{n_{M-1}+n_M+1} \\
&\quad \times \left[1 - \frac{P_{M-1}}{P_{M-2}}\right] \left[\frac{P_{M-2}}{P_{M-3}}\right]^{n_{M-2}+n_{M-1}+n_M+2} \cdots \\
&\quad \times \left[\frac{P_2}{P_1}\right]^{n_2+n_3+\cdots+n_M+M-2} \left[1 - \frac{P_2}{P_1}\right] \left[\frac{P_1}{P_0}\right]^{n_1+n_2+\cdots+n_M+M-1} \left[1 - \frac{P_1}{P_0}\right]
\end{aligned}$$

ここで、 n_{M-1}, \dots, n_1 は図6に示す通り、各観測区間 $[e_M, e_{M-1}], \dots, [e_2, e_1]$ に入る打ち切られたデータ数である ($y_i > 0$ となるデータ数は M 個)。最初の式では各 $P_i (i = 1, \dots, M-1)$ が隣接する3項に現れている。この段階では微分に拠らず次式で各 $P_i (i = 0, \dots, M-1)$ を割り込み (と掛け合わせで相殺し)、2項に隣接する比率 $P_i/P_{i-1} (i = 1, \dots, M)$ で表される尤度関数に帰着される。そこで $\operatorname{argmax}_x x^\alpha(1-x) = \alpha/(1+\alpha)$ を用いれば次の解が得られる。

$$\begin{aligned}
\frac{P_1}{P_0} &= \frac{n_1 + n_2 + \cdots + n_M + M - 1}{n_1 + n_2 + \cdots + n_M + M}, \\
\frac{P_2}{P_1} &= \frac{n_2 + \cdots + n_M + M - 2}{n_2 + \cdots + n_M + M - 1}, \\
&\dots \\
\frac{P_M}{P_{M-1}} &= \frac{n_M}{n_M + 1},
\end{aligned} \tag{15}$$

そこで上式と $P_0 \equiv 1$ より、 $P_1 = \frac{P_1}{P_0} = \frac{n_1+n_2+\cdots+n_M+M-1}{n_1+n_2+\cdots+n_M+M}$ と繰り返せばすべての $P_i (i = 0, \dots, M)$ をノンパラメトリックに推定することができる。

2.3 セミパラメトリック・トービットモデル

Kaplan=Meier 推定量 \hat{F} を用いることで、誤差分布関数 F を特定化せずとも、傾き β については推定が可能になる。3つのセミパラメトリック・モデルを紹介するにあたり、標準的な、つまり $u_i \sim N(0, \sigma^2)$ を仮定したパラメトリック・モデルとの対応関係は次の通りである。

Semiparametric Heckman 法¹⁵ \leftrightarrow Heckman 推計 (11) 式

¹⁵Horowitz (1986) ではすべてのデータ ($y_i = 0$ を含む) を用いた場合のみを対象とし

Horowitz 法¹⁶ ↔ Heckman 推計 (13) 式
 Buckley=James 法¹⁷ ↔ EM アルゴリズム (1.3 節)
 以下、各々の目的関数と推計手続の概略を述べる。

2.3.1 Semiparameric Heckman 法

まず、(11) 式の導出に対応して、

$$E(y|y > 0) = x'\beta + E(u|u > -x'\beta) \quad (16)$$

$$E(u|u > x'\beta) = \int_{-x'\beta}^{\infty} uf(u)du / [1 - F(-x'\beta)] \quad (17)$$

F の関数として表すために、部分積分を用いて、

$$\begin{aligned} \int_{-x'\beta}^{\infty} uf(u)du &= [u(F-1)]_{-x'\beta}^{\infty} + \int_{-x'\beta}^{\infty} (1-F)du \\ &= x'\beta[F(-x'\beta) - 1] + \int_{-x'\beta}^{\infty} (1-F)du \end{aligned} \quad (18)$$

よって、

$$E(y|y > 0) = \int_{-x'\beta}^{\infty} (1-F)du / [1 - F(-x'\beta)] \quad (19)$$

$$\text{Minimize}_b \sum_{y_i > 0} \left[y_i - \frac{\int_{-x'_i b}^{\infty} (1 - \hat{F}_b) du}{1 - \hat{F}_b(-x'_i b)} \right]^2 \quad (20)$$

ここで \hat{F}_b は、 β の推定値 b で評価した Kaplan=Meier 推定量である。 \hat{F}_b が階段関数なので、例えば $\beta \in [0, 2]$ と仮定して、 b を逐次 (0.00, 0.01, ..., 1.99, 2.00) において評価 (grid search) すればよいであろう。

ているが、その方が目的関数が幾分簡単なようである。しかし、パラメトリック・モデルにおいて部分データ ($y_i > 0$ のみ) を用いた場合との優劣は明らかでなかったことに注意する必要がある。

¹⁶Horowitz (1986), "A distribution-free least squares estimator for censored linear regression models," *Journal of Econometrics*, Vol. 32, 59-84.

¹⁷Buckley=James (1979), "Linear regression with censored data," *Biometrika*, Vol. 66, 429-436.

2.3.2 Horowitz 法

(19)式を用いて、

$$Ey = [1 - F(-x'\beta)]x'\beta + \int_{-x'\beta}^{\infty} uf(u)du = \int_{-x'\beta}^{\infty} (1 - F)du \quad (21)$$

$$\text{Minimize } \sum_b \left[y_i - \int_{-x'b}^{\infty} (1 - \hat{F}_b)du \right]^2 \quad (22)$$

2.3.3 Buckley=James 法

(ステップ1) 初期推定量 $\hat{\beta}_{r-1} (r = 1, 2, \dots)$ を与え、Kaplan=Meier 推定量 $\hat{F}_{\hat{\beta}_r}$ を求める。

(17)-(19) 式の導出と同様の議論で、下記の (27) 式を得ることが解る。

$$E(y|y < 0) = x'\beta + E(u|u < -x'\beta) \quad (23)$$

$$E(u|u < -x'\beta) = \int_{-\infty}^{-x'\beta} uf(u)du / F(-x'\beta) \quad (24)$$

$$\int_{-\infty}^{-x'\beta} uf(u)du = [uF]_{-\infty}^{-x'\beta} - \int_{-\infty}^{-x'\beta} Fdu = -x'\beta F(-x'\beta) - \int_{-\infty}^{-x'\beta} Fdu \quad (25)$$

$$E(y|y < 0) = - \int_{-\infty}^{-x'\beta} Fdu / F(-x'\beta) \quad (26)$$

(ステップ2) $y_i = 0$ となる観測値に対しては、 $y_i^0 = - \int_{-\infty}^{-x'\hat{\beta}_1} \hat{F}du / \hat{F}(-x'\hat{\beta}_1)$ と、(27) 式に置き換える (E-step に対応)。

(ステップ3) 最小二乗推定量 (OLSE) $\hat{\beta}_2$ を求める (M-step に対応)。

そして、step-1.2.3. を繰り返し β_r を収斂させる。

ここで説明した3つのモデルに Powell の LAD 法¹⁸ (Least absolute deviations estimator) を加え、各セミパラメトリック・トービットモデルを

¹⁸Powell の LAD 法は $\text{argmin}_{\beta} \sum_i |y_i - \max\{0, x_i'\beta\}|$ で与えられるが、この方法は Kaplan=Meier 推定量を用いない。この推定法が頑強性を持つ直観的な解釈は、i.i.d. サンプルの場合を考えると、中位数は (平均より左の) 打ち切りに対して影響を受けないが、平均値はその限りではないことである。

比較したモンテカルロ実験は、Choon-Geol Moon(1989)¹⁹ に詳しい。同時に Choon-Geol Moon (1989) では、約30~40%の打ち切りを伴うエンゲル曲線 (movie/ transportation expenditure) の推計がなされている。標準的な Tobit MLE の残差に拠る Kaplan=Meier 推定量からは、誤差分布が両側指数分布 (Laplace distribution) に近い可能性が示唆されており、セミパラメトリック・トービット推計が有用となる実証例である。

¹⁹Choon-Geol Moon (1989) "A Monte Carlo Comparison of Semiparametric Tobit Estimators," *Journal of Applied Econometrics*, Vol. 4, 361-382.

第3回：一般化トービット (Generalized Tobit) モデル

2005年5月18日

概要

Amemiya et. al.(1993)¹では家計のポートフォリオ選択を考慮した投資行動の分析が行われている。ポートフォリオ自体の選択は離散的 (discrete) な変数の選択とみなし、所与のポートフォリオの下での投資額は連続的 (continuous) な変数とみなす。このようなモデルの分析枠組みは離散・連続選択モデルとも呼ばれるモデルであり、いくつかの推定手法が存在する。ここでは Amemiya et al.(1993) に沿ってそれらの推定手法を説明する。

1 一般化トービット (Generalized Tobit) モデルとは?

一般化トービット (Generalized Tobit) モデル、あるいは離散・連続選択モデル (discrete/continuous choice model) は離散的な選択と連続的な選択が混在する計量経済モデルを意味している。この種の統計的モデルを推定する問題は計量経済学における分析対象においては必ずしもすべての離散的な選択行動に関係する連続変数を観測できるわけではないということから生じる。例えば、Dubin=MacFadden (1984)²では、家計の電力需要を分析する際に Amemiya (1985) がタイプ5のトービット (Tobit) モデルと名付けた統計モデルをさらに多項目へ拡張した多項 (multinomial)

¹Amemiya, T., M. Saito and K. Shimono (1993), "A Study of Household Investment Patterns in Japan: An Application of Generalized Tobit Model," *The Economic Studies Quarterly*, Vol. 44, pp. 13-28.

²Dubin, J. A. and D. MacFadden (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, Vol. 52, pp. 345-362.

モデルを用いている。彼らの分析では特に長期の電力需要を考える場合、家計が電気製品の組み合わせを調整することによる効果を明示的に考慮する必要があった。電気製品の組み合わせは離散的な選択であり、電力需要は連続的な選択である。したがって、電気製品の選択行動も含めて電力の需要を分析する場合、離散・連続選択モデルを用いることとなった。しかしながら、実際には各個人についてすべての電気製品の組み合わせの下での電力消費が観測されるわけではなく、最適化された選択下での結果のみが観測されるため、単純な推定により意味のある結果を得ることはできない。またこの種の初期の計量経済学的研究としては Duncan (1980)³ があるが、工場の立地選択を考慮した生産関数モデルとその推定方法が提案されている。先の例と同様に、すべての立地における投入・生産量が観測されるわけではなく、最適な立地の選択が行われた結果のみが観測されることに注意する必要がある。第3の例として、Amemiya et al. (1993) を挙げておくと、彼らの研究では家計の投資行動を分析する際のモデルとしてどのようなポートフォリオを組むかという、ポートフォリオ選択を考慮した投資量決定の計量モデルが用いられている。そこで分析された統計モデルは、家計はまず預貯金、債券、株式の3資産をどのように持つかというポートフォリオ選択を行い、その後それぞれの資産への投資額を決定するという計量経済モデルである。この場合においてもすべてのポートフォリオ選択における投資額が観測されるわけではなく、最適な選択がなされた結果が観測されていると考えられる。

以下ではこのような離散・連続選択モデルについて3つの推定手法を説明する。それぞれDubin=MacFadden (1984) の手法、Amemiya et al. (1993) で提案された同時方程式トービット・モデル (simultaneous equation Tobit model) を用いた手法と逐次トービット・モデル (sequential Tobit model) を用いた手法である。

2 Dubin=MacFadden (1984) の推定法

Dubin=MacFadden(1984) では、連邦エネルギー庁(現エネルギー省)による1975年における調査から家計の電力需要データを用いて、電気器具の組み合わせを考慮した家庭用の電力需要の分析を行っている。まず間

³Duncan, G. (1980), "Formulation and Statistical Analysis of the Mixed Continuous/Discrete Variable Model in Classical Production Theory," *Econometrica*, Vol. 48, pp. 839-852.

接効用 (indirect utility) 関数を以下のように定式化する。

$$U_{ji} = \left[\alpha_0^j + \frac{\alpha_1}{\beta} + \alpha_1 p_{1i} + \alpha_2 p_{2i} + w_i' \gamma + \beta (y_i - r_{ji}) + \eta_i \right] e^{-\beta p_{1i}} + \epsilon_{ji} \quad (1)$$

ここで下付きの i は i 番目の家計、 U_{ji} は i 番目の家計が j タイプの電気器具を選択した場合の間接効用関数、 p_{1i} と p_{2i} はそれぞれ電気とガスの料金をそれぞれ表している。また y_i は所得、 r_{ji} は j タイプの電気器具を選択した場合のコスト、 w_i は家計に特有の変数をあらわす。 w_i には世帯人数や気候を表す変数などが含まれている。誤差項は η_i と ϵ_{ji} であり、相関を持つと仮定する。そして家計は最も高い間接効用を与えるような電気器具の組み合わせを選択し、その下で電気を消費すると考える。

電力需要はロイの恒等式⁴ $-\frac{\partial U/\partial p_1}{\partial U/\partial y}$ より

$$x_{ji} = \alpha_0^j + \frac{\alpha_1}{\beta} + \alpha_1 p_{1i} + \alpha_2 p_{2i} + w_i' \gamma + \beta (y_i - r_{ji}) + \eta_i \quad (2)$$

となる。消費額 x_{ji} と選択の決定に伴う誤差に相関があると見なすことが自然であり η_i と ϵ_{ji} が相関し、誤差項 ϵ_{ji} がグンベル分布に従うとする。さらに η と ϵ の関係が線型だと仮定すると、

$$\eta_i = \frac{\sqrt{6}\sigma}{\pi} \sum_j R_j (\epsilon_{ji} - r) + \xi_i \quad (3)$$

と書ける。更に $\sum_j R_j = 0$, $\sum_j T_j^2 < 1$ 、確率変数 ξ_i は ϵ_{ji} と独立であると仮定すると $E(\eta_i) = 0, V(\eta) = \sigma^2$ と表すことができる。(付論を参照。) ここで i 番目の家計が j タイプの組み合わせを選んだ場合に 1 をとるようなダミー変数 D_{ji} を用いると、推定式は

$$\begin{aligned} \sum_j D_{ji} x_{ji} &= \sum_j D_{ji} \alpha_0^j + \frac{\alpha_1}{\beta} + \alpha_1 p_{1i} + \alpha_2 p_{2i} + w_i' \gamma + \beta \left(y_i - \sum_j D_{ji} r_{ji} \right) + \eta_i \\ &\equiv \alpha' s_i + \beta' \sum_j D_{ji} z_{ji} + \eta_i \end{aligned} \quad (4)$$

⁴価格ベクトル p と所得 y を引数とする間接効用関数 V と i 財への需要 x_i に関して恒等的に $-\frac{\partial V(p,y)/\partial p_i}{\partial V(p,y)/\partial y} = x_i(p,y)$ が成立する。

とかける。ここで

$$\alpha' s_i = \begin{bmatrix} \frac{\alpha_1}{\beta}, & \alpha_1, & \alpha_2, & \gamma', & \beta \end{bmatrix} \begin{bmatrix} 1 \\ p_{1i} \\ p_{2i} \\ w_i \\ y_i \end{bmatrix} \quad (5)$$

$$\beta' \sum_j D_{ji} z_{ji} = \begin{bmatrix} \alpha_0^1, & \alpha_0^2, & \dots, & \alpha_0^J, & -\beta \end{bmatrix} \sum_j D_{ji} \begin{bmatrix} e_j \\ r_{ji} \end{bmatrix} \quad (6)$$

と定義した。ただし $e_j = (0, \dots, 1, 0, \dots, 0)'$ は $J \times 1$ のベクトルで、 j 行のみ 1、その他の要素は 0 と定める。

さて、推定式には誤差項として i に依存する η_i しかあらわれていないが、ダミー変数 D_{ji} を通じて ϵ_{ji} が影響し、その結果 $\sum_j D_{ji} z_{ji}$ と η_i が相関を持つ。したがって、単純な最小二乗法により s_i と $\sum_j D_{ji} z_{ji}$ を $\sum_j D_{ji} x_{ji}$ に回帰して得られる推定値には内生性によるバイアスが生じることになる。そこで Dubin=MacFadden (1984) では $j = 1, 2, 3$ の簡単な場合を考えると、

$$\begin{aligned} \lambda_i &= \sum_j D_{ji} E(\eta_i | D_{ji} = 1) \\ &= \frac{\sqrt{6}\sigma}{\pi} [G_{3i} + R_1(G_{1i} - G_{3i}) + R_2(G_{2i} - G_{3i})] \end{aligned} \quad (7)$$

という表現を利用してこの内生性の問題を解決している。(ここで用いる λ_i の性質及び式 (7) の導出は付論 A.1 を参照) すなわちまず推定式を

$$\sum_j D_{ji} x_{ji} = \alpha' s_i + \beta' \sum_j D_{ji} z_{ji} + \lambda_i + \eta_i - \lambda_i \quad (8)$$

と変形する。このとき確率変数 $\eta_i - \lambda_i$ が $\sum_j D_{ji} z_{ji}$ や λ_i と相関を持たないことを利用して方程式を推定することができる。より具体的にはこの定式化による推定は次の 2 段階により行われる。(ただし、 $j = 1, 2, 3$ の場合であることに注意しておく。) ここで間接効用関数は $\delta_i = e^{-\gamma p_{1i}}$ とおくと

$$U_{ji} = (\alpha' s_i + \beta' z_{ji}) \delta_i + \epsilon_{ji}$$

であるので、

$$P_{ji} \equiv \Pr(i \text{ 番目の家計がタイプ } j \text{ を選ぶ}) \\ = \frac{\exp(\beta' z_{ji} \delta_i)}{\exp(\beta' z_{1i} \delta_i) + \exp(\beta' z_{2i} \delta_i) + \exp(\beta' z_{3i} \delta_i)} \quad (9)$$

となる。ここで δ_i が非線形関数であるので、これは非線形多項ロジット (nonlinear multinomial logit) モデルになる。そこで二段階からなる統計的手続き、

1. 尤度 $\prod_i \prod_j P_{ji}^{D_{ji}}$ を最大化して β と γ を推定する。推定された値を P_{ji} に代入して \hat{P}_{ji} 、さらに \hat{G}_{1i} 、 \hat{G}_{2i} と \hat{G}_{3i} を求める。
2. $\sum_j D_{ji} x_{ji}$ を s_i 、 $\sum_j D_{ji} z_{ji}$ 、 \hat{G}_{3i} 、 $\hat{G}_{1i} - \hat{G}_{3i}$ と $\hat{G}_{2i} - \hat{G}_{3i}$ に最小二乗回帰して母数 α 、 β と $\frac{\sqrt{6}\sigma}{\pi} (1, R_1, R_2)$ を推定する。(式 (8) を参照)

により関心のある母数を推定することができる。

3 同時方程式トービット・モデルによる推定法

同時方程式トービット (simultaneous equation Tobit) モデルは Amemiya (1974) により提案され、Amemiya et al. (1993) 実証分析に応用されている。ここではこのモデルをポートフォリオ分析に応用した Amemiya et al. (1993) の実証的内容について説明し、同時方程式トービット・モデルの推定方法について議論する。

Amemiya et al. (1993) では家計の投資行動を資産選択を考慮して分析を行っているが、より具体的には家計は以下の3種類の資産を選択できるとしている。

- 銀行預金・貯金 (y_1)
- 債券 (y_2)
- 株式 (y_3)

ここで経験的に妥当な仮定として、家計は常に預貯金は正の値で持つとしている。(実際、彼らが利用したデータはこの条件が満たされている。) したがって、資産選択として以下の4パターンがあることになるが、変数 y_2 と y_3 はゼロ及び正値をとる変数である。

	y_1	y_2	y_3
1	+	+	+
2	+	+	0
3	+	0	+
4	+	0	0

こうした設定の下で家計の効用最大化問題を定式化する。家計は所与の価格 $\bar{p} = (p_1, p_2, p_3)'$ と所得 $\bar{y} = (y_1, y_2, y_3)'$ の下で、

$$\max U(\bar{y}) = \bar{\epsilon}'\bar{y} + \frac{1}{2}\bar{y}'\bar{A}\bar{y} \quad \text{subject to } \bar{p}'\bar{y} = 1 \quad (10)$$

なる制約つき最大化問題を解くことを考えよう。ただしここでは効用関数は2次効用関数 (quadratic utility function) を仮定していることに注意する。ここで行列とベクトルをそれぞれ

$$\bar{A} = \begin{bmatrix} a_{11} & a' \\ a & A \end{bmatrix} \quad \bar{y} = \begin{bmatrix} y_1 \\ y \end{bmatrix} \quad \bar{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon \end{bmatrix} \quad \bar{p} = \begin{bmatrix} p_1 \\ p \end{bmatrix} \quad (11)$$

と分割して考える。

条件 $y_1 > 0$ とした下で、最大化問題 (10) の解は以下の式の解で与えられる。

$$\left(p_1 A - p a' - a p' + \frac{a_{11}}{p_1} p p' \right) y + p_1 \epsilon - \epsilon_1 p + a - \frac{a_{11}}{p_1} p \leq 0 \leq y \quad (12)$$

ただし、この式はもし最初の不等号が $<$ で成立していれば、2番目の不等号は $=$ になり、一方最初の不等号が $=$ で成立していれば、2番目の不等号は $<$ となるという意味としておく。ここで y に関係しない項を、適当な外生変数 x と誤差項 u の線形結合で置き換えれば

$$-\Gamma y + Bx + u \leq 0 \leq y \quad (13)$$

と書ける。ここで上のように表現できる為には \bar{A} が個人に依存しないと仮定する必要がある。 $\bar{\epsilon}$ は個人に依存してもよい。なお、総資産を M 、4つの1を要素とするベクトルを l であらわすと制約は $l'\bar{y} = M$ となる。両辺を M で割れば $\bar{p}'\bar{y} = 1$ となる。

このように表現される統計的モデルは同時方程式トービット・モデルと呼ばれている。この同時方程式トービット・モデルについては Amemiya

(1974)⁵ が導入した。その後、Ransom (1987)⁶ が効用関数の最大化における Kuhn=Tucker 条件より導けることを示したので、経済学的にも意味がある解釈が可能であることが明瞭となった。

ここで (13) が y について解を持つための必要十分条件は Γ の主小行列が正である。(厳密な証明は Amemiya(1974) の Theorem 3 を参照。) この条件は整合条件 (**coherence condition**) と呼ばれている。ここでより単純化した 2 方程式の例を考えると (13) は

$$\begin{aligned} y_1 &\geq \gamma_1 y_2 + w_1 \quad \text{iff } y_1 \geq 0 \\ y_2 &\geq \gamma_2 y_1 + w_2 \quad \text{iff } y_2 \geq 0 \end{aligned} \tag{14}$$

となる。ここで $w = Bx + u$ とおいた。すなわち、

$$\Gamma = \begin{pmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{pmatrix}$$

である。この場合、整合条件は

$$1 - \gamma_1 \gamma_2 > 0 \tag{15}$$

であるが、 $\gamma_1 > 0, \gamma_2 > 0$ のとき Coherency 条件が満たされている場合と満たされていない場合の解の存在範囲を $w_1 - w_2$ 平面上に表せば、図 1 と図 2 のようになる。すなわち Coherency 条件が満たされている場合は解の存在範囲が $w_1 - w_2$ 平面を重ならず分割していることがわかる。Coherency が満たされることの経済学的意味は、 \bar{A} が負値定符号は $\|y\| \rightarrow \infty$ のときに逓減的な効用関数を仮定することを意味している。

なお同時方程式トービット・モデルの尤度関数は Amemiya (1974) で導出されているので、それを用いて最尤推定を行うことができる。

4 逐次トービット・モデルによる推定法

Amemiya et. al. (1993) で議論している統計モデルを逐次トービット (sequential Tobit) モデルと見なし、両側トービット・モデル (two-limit Tobit model) を逐次的に推定することも可能である。そこでこの方法を

⁵Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated," *Econometrica*, Vol. 42, pp. 999-1012.

⁶Ransom, M. (1987), "A Comment on Consumer Demand System with Binding Nonnegative Constraints," *Journal of Econometrics*, Vol. 34, pp. 355-359.

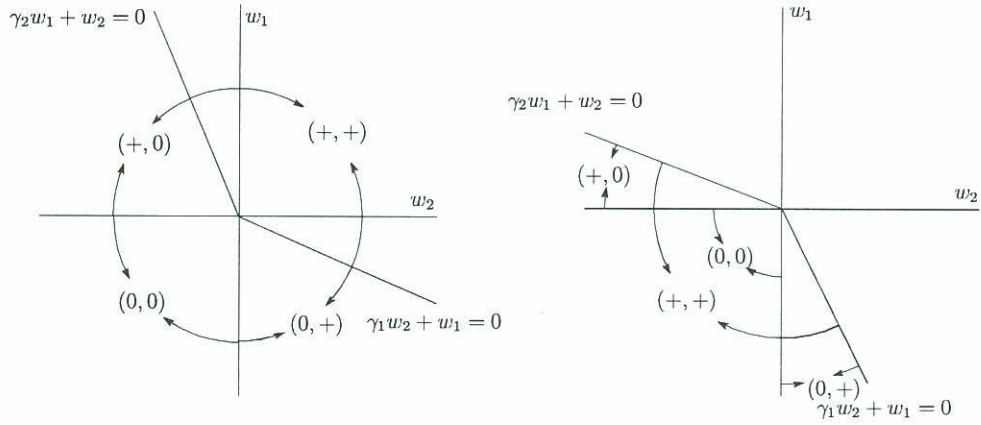


図 1: $\gamma_1 > 0, \gamma_2 > 0, 1 - \gamma_1\gamma_2 > 0$ 図 2: $\gamma_1 > 0, \gamma_2 > 0, 1 - \gamma_1\gamma_2 < 0$

Amemiya et. al.(1993) に従い 3 資産選択のモデルで説明しよう。この場合逐次トービット・モデルとは以下の両側トービット・モデルを 2 段階で推定することである。第 1 段階では総所得 M に対して desired investment 量 y_1^* により決まるトービット方程式、次に残りの所得 $M - y_1$ に対してトービット方程式を導入すると、最後に残差として y_3 を決定するという統計モデルが考えられる。すなわち

$$\text{第 1 段階: } y_1 = \begin{cases} M & \text{if } y_1^* > M \\ y_1^* & \text{if } 0 < y_1^* < M \\ 0 & \text{if } y_1^* < 0 \end{cases} \quad (16)$$

$$\text{第 2 段階: } y_2 = \begin{cases} M - y_1 & \text{if } y_2^* > M - y_1 \\ y_2^* & \text{if } 0 < y_2^* < M - y_1 \\ 0 & \text{if } y_2^* < 0 \end{cases} \quad (17)$$

$$y_3 = M - y_1 - y_2 \quad (18)$$

で与えられる。ただし、ここで登場する潜在変数 y_1^* と y_2^* は独立にそれぞれ $N(x_1'\beta_1, \sigma_1^2)$ と $N(x_2'\beta_2, \sigma_2^2)$ 従うと仮定してトービット分析を行えばよい。

この逐次トービット・モデルは直観的で簡単なモデルである反面、逐次推定を行う順序によって推定結果が異なってしまう可能性があるという問題点が存在する。

5 Amemiya et al. (1993) の分析

Amemiya et al.(1993) では、同時方程式トービット・モデルと逐次トービット・モデルによる推定に加えて、Dubin=MacFadden (1984) タイプの推定手法も紹介している。家計が選択できるポートフォリオは4つあった。あるポートフォリオを選択したという下で最大効用を与える投資額の組をそれぞれ (y_1^1, y_2^1, y_3^1) 、 $(y_1^2, y_2^2, 0)$ 、 $(y_1^3, 0, y_3^3)$ 、 $(M, 0, 0)$ とする。そして

$$y_k^j = x' \beta_k^j + u_k^j \quad (19)$$

と線形関数で近似できるものとする。さらにこのときの間接効用 V_j は、先の y_k^j を効用関数に代入することで求められるが

$$V_j = x' \delta_j + v_j \quad (20)$$

と線形近似できると仮定しよう。

次に誤差項については次のように考えよう。まず $\{v_j\}$ は独立に累積分布関数 $F(v) = e^{-e^{-v}}$ に従うとする。その結果ポートフォリオの選択は多項ロジット・モデルになる。また

$$E(u_k^j | v_1, v_2, v_3) = \frac{\sqrt{6}\sigma_k^j}{\pi} \sum_{i=1}^3 R_k^j(i)(v_i - \gamma) \quad (21)$$

$$V(u_k^j | v_1, v_2, v_3) = (\sigma_k^j)^2 \left[1 - \sum_{i=1}^3 R_k^j(i)^2 \right] \quad (22)$$

とする。ここで $\gamma \cong 0.577$ (オイラーの定数(Euler's constant))、 $\sum_{i=1}^3 R_k^j(i) = 0$ 、 $\sum_{i=1}^3 R_k^j(i)^2 < 1$ である。

この実証分析ではまず Dubin=MacFadden (1984) 同様に2段階推定を行っている。すなわち、次のような推定を行っている。

1. まず

$$E(v_i | j \text{ が選ばれた}) = \begin{cases} \gamma - \log P_i & \text{if } i = j \\ \gamma + \frac{P_i}{1-P_i} \log P_i & \text{if } i \neq j \end{cases} \quad (23)$$

に多項ロジット・モデルで推定した P_i を代入して、 $E(v_i | j \text{ が選ばれた})$ を推定する。

2. 続いて

$$\begin{aligned}
 E(y_k^j | j \text{ が選ばれた}) &= x' \beta_k^j + E(u_k^j | j \text{ が選ばれた}) \\
 &= x' \beta_k^j + \frac{\sqrt{6} \sigma_k^j}{\pi} \sum_{i=1}^3 R_k^j(i) [E(v_i | j \text{ が選ばれた}) - \gamma]
 \end{aligned}
 \tag{24}$$

となっているので、第一段階で求めた $E(v_i | j \text{ が選ばれた})$ を代入して、最小二乗法で推定する。

これまで述べた推定方法について比較してみよう。Amemiya et al.(1993)では上で述べた3つの離散・連続モデルの推定手法を日本の家計の投資行動のデータを用いて比較している。まず推定値が満たすべき符号条件を満たしているかどうかであるが、おおむねどの推定手法で推定しても、所得、資産、持ち家などの係数の符号条件は満たされている。ここで、特に

$$\frac{\partial E(y)}{\partial x_i}$$

の係数が実証的には重要であることに注意しよう。また予測の観点から、推定に用いなかったサンプルの標準誤差を求め比較を行った。結果は、表

表 1: 標準誤差の比較 (Amemiya et al.(1993) Table 14)

	long-term income-gain assets	capital-gain assets
OLS	116.2	188.5
同時方程式トービット	111.4	182.1
Dubin=MacFadden(1984) タイプ	113.0	183.1
逐次トービット	108.6	177.6

1にまとめられており、トービットタイプの推定手法が最小二乗法に比べて優れており、また Dubin=MacFadden (1984) タイプの推定手法は同時方程式トービット・モデルによる推定手法と同程度に優れていることがわかった。

A 付論

A.1 $\lambda_i = \sum_j D_{ji} E(\eta_i | D_{ji} = 1)$ の性質および式(7)の導出

まず $E\lambda_i = 0$ と $E \sum_j D_{ji} z_{ji} (\eta_i - \lambda_i) = 0$ であることを示す。単純化のため i を落とす。

1. $E\lambda = 0$ を示す。まず $E\eta = 0$ である。また $D_j = 1$ のとき $Y = j$ なる変数 Y を定義すると、

$$\begin{aligned} E\eta &= E_Y E(\eta | Y) = \sum_j \Pr(Y = j) E(\eta | Y = j) \\ &= \sum_j \Pr(D_j = 1) E(\eta | D_j = 1) = E\lambda \end{aligned}$$

よって $E\lambda = 0$ 。

2. すべての k に対して $ED_k(\eta - \lambda) = 0$ を示すことで、 $E \sum_j D_{jz_j} (\eta - \lambda) = 0$ を示す。まず $ED_k\lambda = E_{D_k} D_k E(\eta | D_k = 1)$ 。また $ED_k\eta = E_{D_k} D_k E(\eta | D_k)$ 。ここで

$$E(\eta | D_k) = D_k E(\eta | D_k = 1) + (1 - D_k) E(\eta | D_k = 0) \quad (25)$$

と分解できることを用いると、 $E_{D_k} D_k E(\eta | D_k) = E_{D_k} D_k E(\eta | D_k = 1)$ 。よって $ED_k\lambda = ED_k\eta$ であり、すべての k に対して $ED_k(\eta - \lambda) = 0$ 。

次に式(7)を導出する。まず $F(\epsilon_{ji})e^{-e^{-\epsilon}}$ と仮定する。(ここで $E\epsilon_{ji} = \gamma \cong 0.5777^7$ 、 $V\epsilon_{ji} = \frac{\pi^2}{6}$ である。) また、 ξ_i を $\{\epsilon_{ji}\}$ と独立な $E(\xi_i) = 0$ かつ $V(\xi_i) = \sigma^2 (1 - \sum_j R_j^2)$ となる確率変数とする。ここで $\sum_j R_j = 0$ 、 $\sum_j R_j^2 < 1$ を満たしている。さらに、 $\eta_i = \frac{\sqrt{6}\sigma}{\pi} \sum_j R_j (\epsilon_{ji} - \gamma) + \xi_i$ であると仮定する ($E(\eta) = 0$ 、 $V(\eta) = \sigma^2$)。さらに単純化の為に $j = 1, 2, 3$ とすると、付論(A.2)の結果を用いて

$$\begin{aligned} E(\eta_i | D_{1i} = 1) &= \frac{\sqrt{6}\sigma}{\pi} \sum_j R_j E(\epsilon_{ji} | D_{1i} = 1) \\ &= \frac{\sqrt{6}\sigma}{\pi} \left[-R_1 \log P_{1i} + R_2 \frac{P_{2i}}{1 - P_{2i}} \log P_{2i} + R_3 \frac{P_{3i}}{1 - P_{3i}} \log P_{3i} \right] \end{aligned} \quad (26)$$

⁷オイラーの定数。

よって

$$\begin{aligned}
\lambda_i &= \frac{\sqrt{6}\sigma}{\pi} \left\{ D_{1i} \left[-R_1 \log P_{1i} + R_2 \frac{P_{2i}}{1-P_{2i}} \log P_{2i} + R_3 \frac{P_{3i}}{1-P_{3i}} \log P_{3i} \right] \right. \\
&\quad + D_{2i} \left[-R_2 \log P_{2i} + R_1 \frac{P_{1i}}{1-P_{1i}} \log P_{1i} + R_3 \frac{P_{3i}}{1-P_{3i}} \log P_{3i} \right] \\
&\quad \left. + D_{3i} \left[-R_3 \log P_{3i} + R_1 \frac{P_{1i}}{1-P_{1i}} \log P_{1i} + R_2 \frac{P_{2i}}{1-P_{2i}} \log P_{2i} \right] \right\} \\
&= \frac{\sqrt{6}\sigma}{\pi} \left\{ R_1 \left[-D_{1i} \log P_{1i} + (1-D_{1i}) \frac{P_{1i}}{1-P_{1i}} \log P_{1i} \right] \right. \\
&\quad + R_2 \left[-D_{2i} \log P_{2i} + (1-D_{2i}) \frac{P_{2i}}{1-P_{2i}} \log P_{2i} \right] \\
&\quad \left. + R_3 \left[-D_{3i} \log P_{3i} + (1-D_{3i}) \frac{P_{3i}}{1-P_{3i}} \log P_{3i} \right] \right\} \\
&\equiv \frac{\sqrt{6}\sigma}{\pi} (R_1 G_{1i} + R_2 G_{2i} + R_3 G_{3i}) \\
&= \frac{\sqrt{6}\sigma}{\pi} [G_{3i} + R_1 (G_{1i} - G_{3i}) + R_2 (G_{2i} - G_{3i})]
\end{aligned}$$

となる。

A.2 Dubin=MacFadden (1984) の式 (23) の導出

確率変数 ϵ_1 、 ϵ_2 、 ϵ_3 が独立に累積分布関数 $e^{-e^{-\epsilon}}$ にしたかうと仮定しよう。このとき

$$\begin{aligned}
E(\epsilon_1 | 1 \text{ が選ばれる}) &= E(\epsilon_1 | \epsilon_2 < \epsilon_1 + \mu_1 - \mu_2, \epsilon_3 < \epsilon_1 + \mu_1 - \mu_3) \\
&= \frac{1}{P_1} \int_{-\infty}^{\infty} \int_{-\infty}^{\epsilon_1 + \mu_1 - \mu_2} \int_{-\infty}^{\epsilon_1 + \mu_1 - \mu_3} \epsilon_1 f(\epsilon_1) f(\epsilon_2) f(\epsilon_3) d\epsilon_1 d\epsilon_2 d\epsilon_3 \\
&= \frac{1}{P_1} \int_{-\infty}^{\infty} \epsilon_1 e^{-\epsilon_1} e^{-e^{-\epsilon_1}} e^{-e^{-(\epsilon_1 + \mu_1 - \mu_2)}} e^{-e^{-(\epsilon_1 + \mu_1 - \mu_3)}} d\epsilon_1 \\
&= \frac{1}{P_1} \int_{-\infty}^{\infty} \epsilon_1 e^{-\epsilon_1} e^{-\alpha_1 e^{-\epsilon_1}} d\epsilon_1 \quad (\alpha_1 = 1 + e^{\mu_2 - \mu_1} + e^{\mu_3 - \mu_1}) \\
&= -\frac{1}{P_1} \int_0^{\infty} \log x e^{-\alpha_1 x} dx \quad (e^{-\epsilon_1} = x) \\
&= \frac{1}{P_1} \frac{1}{\alpha_1} (\gamma + \log \alpha_1)
\end{aligned} \tag{27}$$

最後の等式は Gradshteyn=Ryzhik(2000)⁸の 4.331(1)、p.537 による。ここで $P_1 = \frac{1}{\alpha_1}$ (Amemiya(1985)⁹ の式 (9.3.53)、p.301 を参照) より、

$$E(\epsilon_1 | 1 \text{ が選ばれる}) = \gamma + \log \alpha_1 = \gamma - \log P_1 \quad (28)$$

となることが知られている。同様に

$$E(\epsilon_2 | 1 \text{ が選ばれる}) = \gamma + \frac{P_2}{1 - P_2} \log P_2 \quad (29)$$

$$E(\epsilon_3 | 1 \text{ が選ばれる}) = \gamma + \frac{P_3}{1 - P_3} \log P_3 \quad (30)$$

が求まる。この計算には $E(\epsilon_2 | 1 \text{ が選ばれる})$ と同様 Gradshteyn=Ryzhik (2000) の 6.224(1)、p.641 及び 8.211(1)、p.925 を用いる。これらの文献ではそれぞれ

$$6.224(1): \int_0^{\infty} E_i(-\beta x) e^{-\mu x} dx = -\frac{1}{\mu} \log \left(1 + \frac{\mu}{\beta} \right), \quad \operatorname{Re}(\beta + \mu) \geq 0, \mu > 0 \quad (31)$$

$$8.211(1): E_i(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt, \quad x < 0 \quad (32)$$

となることが示されている。

⁸Gradshteyn, I. S. and I. M. Ryzhik (2000), *Table of Integrals, Series, and Products*, A. Jeffrey eds., San Diego: Academic Press.

⁹Amemiya, T. (1985), *Advanced Econometrics*, Oxford: Basil Blackwell

第4回 デュレーション(Duration)モデル

2005年5月25日

概要

デュレーション (duration) モデルの基本的分析枠組みを説明し、次に経済学における最適化を考慮した構造的デュレーション (structural duration) 計量モデルを用いた分析法を説明する。労働経済学等で現れる幾つかの計量経済分析も解説する。

1 デュレーション (duration) 分析の基礎

(duration、継続時間や生存時間と呼ばれることがある) モデルを用いた統計的分析は生物統計では人間の生存や医学的治療効果を分析する生存時間解析 (survival analysis)、工学では機械の故障などを扱う寿命の分析、あるいは地震など自然現象の分析など様々な分野で応用されている。経済学においては、例えば労働経済学においては失業期間の分析、あるいはファイナンスにおける株価の秒単位で計測される変動の分析などをはじめ多くの計量分析で用いられている。

例えば労働経済学での失業の分析では雇用の状態を E、失業の状態を U として失業の継続時間 (duration) を分析することが行われているので、この例を取りあげてみる。状態が二つある継続時間モデルは 2 状態モデル (2 state model) と呼ばれ、二つの duration を問題にしていると見なせる。さらに、非労働力 (out of labour force) 状態 O まで考慮すると、3 状態モデル (3 state model) を考える必要も生じる。この場合には例えば失業状態が終わっても、雇用状態か非労働力状態かの二つの状態となる可能性があるので、競争的リスク (competing risk) モデルを使った分析を行う必要性も生じる。

ここでは、まずもっとも簡単に継続時間 (寿命) を表す確率変数 T が i.i.d. となる場合から分析を始めよう。継続時間 (寿命) 変数 (duration variable) を T 、その分布関数を $F(t) = P(T < t)$ とおく。ここでハザード (hazard、危険率) を

$$P(t < T, t + \Delta t | T > t)$$

とすれば、確率変数の密度関数が存在する場合は

$$P(t < T < t + \Delta t | T > t) = \frac{P(t < T < t + \Delta t)}{P(T > t)} \\ \cong \frac{f(t)\Delta t}{1 - F(t)}$$

で与えられる。継続時間の分析では密度関数よりもハザード関数の方が直観的に理解しやすい。と言うのはハザード関数は t 時点より長生きする条件のもとで (Δt の期間の内に) 死亡する確率を意味するので解釈が容易となるからである。

ここで両辺を Δt で割ればハザード関数 $\lambda(t)$ と分布関数の関係は

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \Leftrightarrow F(t) = 1 - \exp\left(-\int_0^t \lambda(s) ds\right)$$

と与えられる。したがって、ハザード関数のモデルを考えることにより色々な確率関数 $F(t)$ を導くことができる。ここで簡単な例を挙げておく。

- 指数分布 (exponential distribution) $\lambda(t) = \lambda$
指数分布はもっとも基本的な分布であるが、ハザード関数は一定となる。したがって、明らかに人間の寿命のモデルとしては不適切であろう¹ 人間の寿命のハザード関数はU字型になると考えられる。
- ワイブル分布 (Weibull distribution) $\lambda(t) = \lambda \alpha t^{\alpha-1}$ $\alpha \leq 1 \Leftrightarrow \partial \lambda / \partial t \leq 0$
指数分布を一般化したのがワイブル分布である。ただし、ワイブル分布はU字型のハザード関数は含まないことに注意しておく。

ここで計量経済分析におけるハザード関数 $\lambda(t)$ モデルとして著名な研究について言及しておこう。Lancaster (1979)² は英国における失業率の duration を説明変数を含んだワイブル分布でモデル化し、

$$\lambda(t) = \exp(x_i' \beta) \alpha t^{\alpha-1}$$

となるモデルを用いて分析を行っている。その研究では説明変数として変数 (log age, log unemployment rate of the area, log replacement -unemployment benefit/earnings form last job) 等を用いたが、これらの説明変数を順次増やしていくと、継続時間の母数 α の推定値が 0 から 1 に次第に近づくことを発見している。この現象を理解する為に、次のような各々 0.5, 0.2, 0.1 の一定のハザードを持つ 3 種類の人々のデータから構成される 3 期間の例を数値的に考察

¹例えば 24 才の健康な院生について $t = 24 \times 365 \times 24$, $\Delta t = 1.5$ とすると、一般に講義中のハザード率は年配の教授の持つハザード率よりはるかに小さいであろう。

²Lancaster, T. (1979) "Econometric Methods for the Duration of Unemployment", *Econometrica*.

して見れば（表1）、説明変数を入れると個体差がとらえられることにより結果として共通としている母数 α の推定値は1に近づく（総和ハザード率が小さくなっていくこと）と考えられよう。ここで例えば0.5のハザードを持つ人は1期のはじめに1000人居たが、その期の終わり（すなわち2期の初め）には500人になることを意味する。Lancasterの実証分析では失業率のdurationについて仮説 $\alpha = 1$ を棄却することができなかった。

労働経済学における考察からは一般的にこの母数 α の値について事前的にはっきりしたことは云えない。失業率が長くなるとより低い賃金でも働くインセンティブが高まると考えれば $\lambda(t)$ は増加関数となる。他方、長く働かない状態が続くと福祉(welfare)の恩恵を受け、働くインセンティブが低くなるなら、時間の減少関数と考えることもできる。Lancasterはハザード率がほぼ一定となることを実証的に証明したわけである。

表 1: Illustrative Example

期間	1		2		3	
ハザード関数	生存	死亡	生存	死亡	生存	死亡
0.5	1000	500	500	250	250	125
0.2	1000	200	800	160	640	128
0.1	1000	100	900	90	810	81
総和	3000	800	2200	500	1700	334
総和ハザード率	0.267		0.227		0.196	

次に Flinn=Heckman (1982)³ はワイブル分布を含む5つの母数を持つモデルを用いて分析している。すなわち、ハザード関数モデルとして

$$\lambda(t) = \exp\left(\gamma_0 + \gamma_1 \frac{t^{\lambda_1 - 1}}{\lambda_1} + \gamma_2 \frac{t^{\lambda_2 - 1}}{\lambda_2}\right)$$

を考察している。ここで、ハザード関数に表れる関数

$$\frac{t^\lambda - 1}{\lambda}$$

はボックス・コックス変換 (Box-Cox transformation) と呼ばれているが、ベキ関数と $\lambda \rightarrow 0$ のとき対数関数 $\log t$ を特殊ケースとして含んでいることに注意する。すなわち、 $\lambda_1 = 0$ かつ $\lambda_2 = 0$ であるならば $\lambda(t) = \exp(Y_0 + t^{\beta_1} + t^{\beta_2})$ となりワイブル分布を特殊な場合として含んでいる。

³Flinn, C. J. and J. J. Heckman (1982) "Models for the Analysis of Labor Force Dynamics", *Advances in Econometrics*.

さらに、Gritz (1993)⁴ が考察したハザード関数は対数ロジスティック・モデルであり、

$$\lambda_0(t) = \rho kt^{k-1}/(1 + \rho t^k)$$

で与えられる。このモデルではハザード関数は

$$k \leq 1 \Rightarrow \frac{\partial \lambda}{\partial t} < 0, \quad k > 1 \Rightarrow \lambda_0(t) \text{ 逆 U 字型}$$

を満たしている。したがって、ハザード関数は逆 U 字型となりうることに注意しておく。

また、Sturm (1991)⁵ は U 字型や逆 U 字型のハザード関数となりうるモデルとして

$$\lambda_0(t) = \lambda \exp(Y_1 t + Y_2 t^2)$$

を考察している。

次に計量生物学で特に著名な統計モデルとして Cox (1972) の比例ハザード・モデル (Proportional Hazard Model)⁶ が知られているが、ハザード関数は

$$\lambda_i(t) = \lambda_0(t) \exp(X_i' \beta)$$

で与えられるので、Lancaster のモデルを含むより一般的なモデルと考えることができる。ここで $\lambda_0(t)$ は basic hazard と呼ばれ、特定な形は想定しない。すなわち $\lambda_0(t)$ はノンパラメトリックな方法である Kaplan-Meier 法により推定し、その他の母数は部分尤度 (partial likelihood) を最大化して推定するのが Cox が提唱した分析方法である。

ここでハザード関数を二つに分離して β は Cox の部分最尤法 (Partial MLE) を利用し $\lambda_0(t)$ は Kaplan-Meier (Kaplan-Meier) 推定量 Kaplan and Meier (1958)⁷ で推定する方法の計量経済学での応用例としては Lehrer (1988)⁸ 等がある。

実際には Cox の PMLE (部分最尤法) は部分尤度関数を β に関して最大化、すなわち

$$PL = \prod_i \frac{\lambda_i(t_i)}{\sum_{j \in R_i} \lambda_j(t_i)}$$

を β に関して最大化することで求める。ここで、例えば経済学における銀行の倒産分析の例で考えると R_i は i 銀行が倒産するまで存在していた銀行の数として適用することができる。

⁴Gritz, A. M. (1993), "The Impact of Training on the Frequency and Duration of Employment", *Journal of Econometrics*, 64 pp439-442.

⁵Sturm, K. (1991). "Reliability and Maintenance in European Nuclear Power Plants: A Structural Analysis of a Controlled Stochastic Process", Ph. D. Thesis, Stanford University.

⁶Cox, D. R. (1972) "Regression Models and Life Tables", *JRSS B*.

⁷Kaplan and Meier (1958) "Nonparametric Estimation from Incomplete Observations", *JASA*.

⁸Lehrer, E. L. (1988) "Determinants of Marital Instability: A Cox Regression Model", *Applied Econometrics*.

Cox の部分尤度 (Partial Likelihood)

ここで簡単な設定の下で Cox の部分尤度 (PL) を説明しておこう。二人の寿命が観察され、寿命の確率変数 T_1 と T_2 で表し、互いに独立であって

$$T_1 \sim (F_1, f_1, \lambda_1),$$

$$T_2 \sim (F_2, f_2, \lambda_2)$$

としよう。このとき二人の中でどちらか一方が t 期に死亡したという条件のもとで第一番目の人が死亡する確率は

$$\begin{aligned} & P(T_1 = t | T_1 \geq t, T_2 \geq t, T_1 = t \text{ or } T_2 = t) \\ &= \frac{P(T_1 = t, T_1 \geq t, T_2 \geq t, T_1 = t \text{ or } T_2 = t)}{P(T_1 \geq t, T_2 \geq t, T_1 = t \text{ or } T_2 = t)} \end{aligned} \quad (1)$$

で表せる。次に N (分子) と D (分母) はそれぞれ

$$N = P(T_1 = t, T_2 \geq t) = P(T_1 = t)P(T_2 \geq t),$$

$$\begin{aligned} D &= P(T_1 = t \text{ or } T_2 = t | T_1 \geq t, T_2 \geq t)P(T_1 \geq t, T_2 \geq t) \\ &\cong [P(T_1 = t | T_1 \geq t, T_2 \geq t) + P(T_2 = t | T_1 \geq t, T_2 \geq t)] P(T_1 \geq t)P(T_2 \geq t) \\ &= [P(T_1 = t | T_1 = t) + P(T_2 = t | T_2 \geq t)] P(T_1 \geq t)P(T_2 \geq t) \end{aligned}$$

となる。ただし、 D における近似の意味は

$$P(T_1 = t \text{ and } T_2 = t | T_1 \geq t, T_2 \geq t)$$

が十分小さいとしたものである。ここで上の N と D の表現よりハザード関数 $\lambda(t)$ の定義に注意すると

$$\frac{N}{D} = \frac{\lambda_1(t)}{\lambda_1(t) + \lambda_2(t)} = \frac{e^{x_1\beta}}{e^{x_1\beta} + e^{x_2\beta}}$$

となる。すなわち (1) が上式のように表せることがわかった。

以上の議論を n 個の確率変数の場合に拡張し、各個体がそれぞれ $t_1 < t_2 < \dots < t_n$ に死亡したと考えれば、最終的には Cox の部分尤度は

$$\text{PLE (部分尤度)} = \frac{\lambda_1}{\lambda_1 + \dots + \lambda_n} \cdot \frac{\lambda_2}{\lambda_2 + \dots + \lambda_n} \dots \frac{\lambda_{n-1}}{\lambda_{n-1} + \lambda_n}$$

で与えられる。

ところでCoxの比例ハザードモデルでは基礎ハザード関数 $\lambda_0(t)$ の推定にはKaplan=Meier法が用いられる。そこでこのK-M法について簡単に言及しておく。

Kaplan-Meier in Duration) 推定量

生存時間(duration)の分析は既に説明したTobitモデルと本質的には同様である。まずもっとも単純なi.i.d.の場合を考え、 t_1 を最初に死亡が観察される時間としよう。もしそれまでに死亡した人がいなければ $n_0 = 0$ である。あるいは t_0 と t_1 の間に死亡した人がいればそれを n_0 とする⁹。つづいて t_1, t_2, \dots, t_M に死亡が生じ、 t_k と t_{k+1} の間に n_k 人($k = 1, 2, \dots, M-1$)が失われたと仮定する(図1を参照)。このときトービットにおけるK-M推定の議論の類推より尤度関数は

$$\begin{aligned} L &= S_0^{n_0} (1 - S_1) S_1^{n_1} (S_1 - S_2) S_2^{n_2} \cdots S_{M-1}^{n_{M-1}} (S_{M-1} - S_M) S_M^{n_M} \\ &= \left(1 - \frac{S_1}{S_0}\right) \left(\frac{S_1}{S_0}\right)^{K_1} \left(1 - \frac{S_2}{S_1}\right) \left(\frac{S_2}{S_1}\right)^{K_2} \cdots \left(1 - \frac{S_M}{S_{M-1}}\right) \left(\frac{S_M}{S_{M-1}}\right)^{K_M} \end{aligned}$$

となる。ただし、

$$\begin{aligned} K_1 &= n_1 + n_2 + \cdots + n_M + M - 1, \\ K_2 &= n_2 + \cdots + n_M + M - 2, \\ &\vdots \\ K_j &= n_j + \cdots + n_M + M - j, \\ &\vdots \\ K_{M-1} &= n_{M-1} + n_M + 1, \\ K_M &= n_M \end{aligned}$$

である。ここで導かれた尤度を構成する比について最大化を考えると最尤推定量を導くことができ、比

$$\hat{S}_1 = \frac{K_1}{K_1 + 1}, \quad \left(\frac{\hat{S}_2}{S_1}\right) = \frac{K_2}{K_2 + 1}, \quad \dots, \quad \left(\frac{\hat{S}_M}{S_{M-1}}\right) = \frac{K_M}{K_M + 1}$$

で与えられることがわかる。この最尤推定量の直観的解釈としては

$$\left(\frac{\hat{S}_j}{S_{j-1}}\right) = \frac{t_j \text{期にまだ生存している人の数}}{t_j \text{手前まで生存している人の数}}$$

⁹例えばスタンフォード大学病院は心臓移植で著名であるが、このような場合には最初の患者が追跡できるの
で $n_0 = 0$ としてもよいが、実際に得られる多くの分析データでは $n_0 > 0$ となる場合が多い。

と見ることができる。

比例ハザード・モデルにおける K-M(Kplan-Meier) 推定

次に比例ハザードモデルにおけるハザード関数 $\lambda_i(t)$ と生存確率 $S_i(t)$ は次のように表現される。

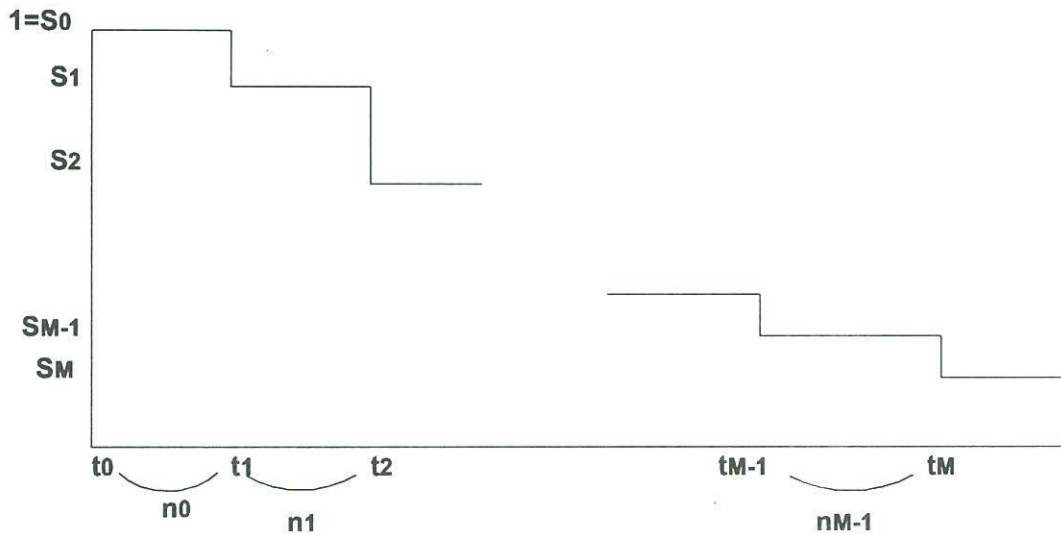


図 1: カプランマイヤー推定量の構成

$$\begin{aligned}
 \lambda_i(t) &= \lambda_0(t)\mu_i \quad (\mu_i = e^{x_i'\beta}), \\
 S_i(t) &= \exp\left[-\int_0^t \lambda_i(\tau)d\tau\right] \\
 &= \exp\left[-\mu_i \int_0^t \lambda_0(\tau)d\tau\right] \\
 &\equiv S_0(t)^{\mu_i}
 \end{aligned}$$

したがって、尤度関数は

$$\begin{aligned}
L &= \prod_0 S_0(t_0)^{\mu_i} [1 - S_0(t_1)^{\mu_1}] \prod_1 S_0(t_1)^{\mu_i} [S_0(t_1)^{\mu_2} - S_0(t_2)^{\mu_2}] \cdots \\
&\quad \times \prod_{M-1} S_0(t_{M-1})^{\mu_i} [S_0(t_{M-1})^{\mu_M} - S_0(t_M)^{\mu_M}] \prod_M S_0(t_M)^{\mu_i} \\
&= (1 - S_1^{\mu_1}) S_1^{\sum_1 \mu_i} (S_1^{\mu_2} - S_2^{\mu_2}) S_2^{\sum_2 \mu_i} \cdots S_{M-1}^{\sum_{M-1} \mu_i} (S_{M-1}^{\mu_M} - S_M^{\mu_M}) S_M^{\sum_M \mu_i} \\
&= \left[1 - \left(\frac{S_1}{S_0} \right)^{\mu_1} \right] \left(\frac{S_1}{S_0} \right)^{K_1} \left[1 - \left(\frac{S_2}{S_1} \right)^{\mu_2} \right] \left(\frac{S_2}{S_1} \right)^{K_2} \cdots \left[1 - \left(\frac{S_M}{S_{M-1}} \right)^{\mu_M} \right] \left(\frac{S_M}{S_{M-1}} \right)^{K_M}
\end{aligned}$$

となる。ただし、記号 \prod_j は時点 t_j で失われたすべての人についての積を意味し、 $S_j = S_0(t_j)$ としている。さらに各 K_i は

$$\begin{aligned}
K_1 &= \sum_1 + \sum_2 + \cdots + \sum_M + \mu_2 + \mu_3 + \cdots + \mu_M, \\
K_2 &= \sum_2 + \cdots + \sum_M + \mu_3 + \cdots + \mu_M, \\
&\vdots \\
K_{M-1} &= \sum_{M-1} + \sum_M + \mu_M, \\
K_M &= \sum_M
\end{aligned}$$

で与えられる。観測値が i.i.d. と見なせる場合と同様にして尤度関数を構成する比について最大化すると、最尤推定量は

$$(\text{MLE}) \left(\frac{S_j}{S_{j-1}} \right)^{\mu_j} = \frac{K_j / \mu_j}{1 + (K_j / \mu_j)}$$

で与えられることがわかる。

競合リスク (Competing Risk) モデル

既に失業の分析例において言及したように、複数のリスク要因 (K 個) を継続時間モデルとして考える必要が生じる場合がある。雇用状態から失業状態になるか、あるいは職業訓練や大学に行くなどの非労働力状態となる可能性を考える場合には異なるリスク要因を介在する。計量生物学の例では T_1 を第1のリスク要因 (脳卒中) での死亡時間、 T_2 を第2のリスク要因 (ガン) での死亡時間、とすると2つの競合リスク・モデルでの分析が必要となる。ここで二つの継続時間変数に対し

$$T = \min(T_1, T_2)$$

であり、 T_1 と T_2 が互いに独立であると仮定すると分析が容易となる。しかしながら、例えば別の例として銀行の倒産確率の分析例などを考えると破産状態になる時間と他行との合併時間と独立でないとする方がより自然であろう。

継続時間変数 T_i と T_j を独立に扱うのは適当でない場合にはまずハザード関数を

$$\lambda_{ik}(t) = \lambda_{0k}(t)e^{\beta'_k x_{ik}(t)} \cdot v_{ik}, \quad k = 1, 2, \dots, K$$

としよう。このとき尤度関数を倒産事象 I、非倒産事象 II に分割し

$$L = \prod_{I:i=1}^{N_1} E \prod_{k=1}^K \lambda_{ik}(t_j)^{\chi(d_j=k)} e^{-\sum_{k=1}^K \int_0^{t_j} \lambda_{ik}(t) dt} \\ \times \prod_{II:i=1}^{N_0} E e^{-\sum_{k=1}^K \int_0^{t_j} \lambda_{ik}(t) dt}$$

として分析を行うことができる。すなわち、ハザード関数にかかわる λ_{ik} のところのみで依存関係が生じると考えれば扱いが容易となるので、実際の実証分析ではこうしたモデルが用いられている。むしろより一般の非独立モデルを考えることはできるが、モデルの識別条件や母数の最尤推定を行うことはより困難となる。

一般の尤度関数 (likelihood function)

ここで打ち切り (censoring) を含みハザード関数が

$$\lambda_i(t) = \lambda_0(t) \exp(x'_{it}\beta)$$

となる場合の尤度関数を考えよう。セミパラメトリック・モデルでは $\lambda_0(t)$ はベースライン・ハザード関数であるが、この関数をノンパラメトリックに推定する必要がある。また、 $\exp(X'_{it}\beta)$ の部分はパラメトリックに推定することが一般的である。まず分布関数を打ち切りがある場合とない場合に分け、(すべての生存区間 t_1 を観測できた場合): $f_1(t_1)$ 、(観測期間より長く生きたばあい): $1 - F_2(t_2)$ (右打ち切り (right-censoring))、とする。例えばスタンフォード病院で 1970~2004 年に心臓移植手術が 3 人に施されそのうち一人はまだ生存していれば、尤度関数は

$$L = f_1(t_1)f_2(t_2)P(T_3 > t_3) \\ = f_1(t_1)f_2(t_2)(1 - F_3(t_3)) \quad (2)$$

となる。より一般的には N_1 が死亡し N_0 人がなお生存している場合には

$$L = \prod_{i=1}^{N_1} f_i(t_i) \prod_{i=1}^{N_0} (1 - F_i(t_i))$$

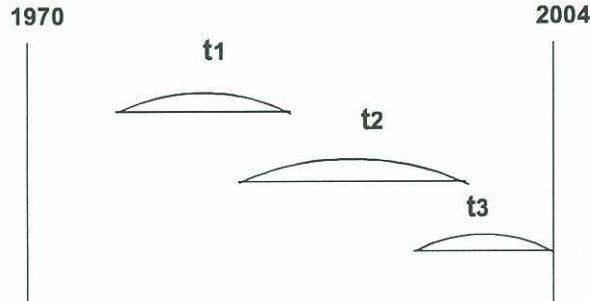


図 2: 右打ち切りデータ：心臓手術の例

で与えられる。

さらに、比例ハザード・モデルを仮定すれば式 (2) は打ち切りがある場合とない場合で区別して

$$L = \prod_{i=1}^{N_1} \lambda_0(t_i) e^{\beta' x_{it}} e^{-\int_0^{t_i} \lambda_0(t) e^{\beta' x_{it}} dt} \times \prod_{i=1}^{N_0} e^{-\int_0^{t_i} \lambda_0(t) e^{\beta' x_{it}} dt}$$

となる。

ここで打ち切りを含む比例ハザード・モデルの困難な点としては尤度関数に表れる積分項

$$\int_0^t \lambda_0(s) \exp(x'_{is} \beta) ds$$

を評価するのは容易でないことが挙げられる。そこで実際には x_{is} を階段関数と仮定し、被積分関数を離散近似して積分を和に書き換えて分析を行っている場合が多い。

次に個体間の (不均一性) (Heterogeneity) の問題を考えよう。例えば

$$\lambda_i(t) = \lambda_0(t) \exp(x'_{it} \beta + v_i)$$

と表現し、誤差項 v の分布を G として分析を行うことが考えられる。確率変数 v_i が与えられたときの i の条件付き確率を $L_i(v_i)$ とおくと尤度関数は

$$LF = \prod_i EL_i(v_i)$$

となるが、ここで期待値 E は分布 G に対して取ったものである。例えば Lancaster (1979) は G にガンマ分布を利用している。これに対して Heckman and Singer (1984)¹⁰ は G を離散分布で近似することにより G も含めた最尤法を提案している。

¹⁰Heckman, J. J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*.

また、時間について離散時間近似 (Discrete-Time Approximation) を行うことも有力な方法として考えられている。 $\lambda_i(t)$ を i 番目の人が時刻 t に死亡した確率と解釈して、例えば議論を単純化して3人の構成員で2期間のモデルを考えてみよう。(例えば Moffitt (1985)¹¹ はハザード関数 $\lambda_i(t) = \lambda_0(t) \exp(x'_{it}\beta)$ のもとで $\lambda_i(t)$ と β を最尤推定している。)

表 2: 3人、2期間のモデル

$i \setminus t$	1	2
1	生存	生存
2	生存	死亡
3	死亡	

この場合、各個人の従う確率はそれぞれ

$$\begin{aligned} & (1 - \lambda_1(1))(1 - \lambda_1(2)), \\ & (1 - \lambda_2(1))\lambda_2(2), \\ & \lambda_3(1) \end{aligned}$$

となるので、互いに独立であるとするれば、尤度関数は

$$\begin{aligned} L = & (1 - \lambda_0(1)e^{x'_{11}\beta})(1 - \lambda_0(2)e^{x'_{12}\beta}) \\ & \times (1 - \lambda_0(1)e^{x'_{21}\beta})\lambda_0(2)e^{x'_{22}\beta} \\ & \times \lambda_0(1)e^{x'_{31}\beta} \end{aligned}$$

で与えられる。

なお、これまで議論した右打ち切り (right-censoring) 問題におけるデータの扱いは、打ち切り確率を考慮すればよいと言う意味で分析上ではそれほど困難な問題を発生させない。これに対して、左打ち切り (left-censoring) データの場合には様々な困難な統計的問題を生じさせるので改めて次回に議論する。

2 構造型デュレーション (structural duration) モデル

これまでに説明してきたのは生存時間分析の一般的な内容とその応用であり、計量経済学の伝統的用語を用いると誘導型 (reduced form) アプローチのモデルと見なすことができる。これ

¹¹Moffitt, R. (1985). Unemployment Insurance and the Distribution of Unemployment Spells. *JOE.*, 28-1.

に対して、より経済学的考察より導かれないわば構造型デュレーション（継続時間）アプローチによる計量分析もかなり盛んに行われているので、次にそうした計量経済学的研究の代表例について議論しよう。

2.1 求職モデル (Search Model)-離散時間 (Discret Time)

経済学的な議論の典型例としては労働経済学における求職 (job search) の議論が重要であるので、まずその分析枠組みを説明し、徐々に複雑な計量経済学的モデルを解説していこう。まず変量を

- λ 雇用機会を得る確率 (Prob(a wage offer arrives)),
- G 賃金の分布関数 (distribution function of wage),
- c 求職コスト (search cost),
- δ 割引率 (discount rate) ($\beta = 1 - \delta$)

により定める。ここでは、計画期間を無限大を仮定し、上記の各変量が期間に依存しない定常的求職 (stationary search) モデルを考えることにする。(これは満期を ∞ とすると分析がかなり容易になり、ベルマン方程式を解くことができるからである。) まず、仮に期間 t で雇用機会を受け入れると賃金流列の現在価値は (一生涯、受け入れた賃金率 $W(t)$ で働くと)、

$$W(t) + \beta W(t) + \beta^2 W(t) \cdots = \frac{1}{1 - \beta} W(t) = \frac{1}{\delta} W(t)$$

で与えられるが、各期間で賃金のオファーが来たときに ($W(t)$ を受け入れるか否かの最適な選択) と、こない場合 (求職コストを投じて来期に高賃金を期待する) を勘案すると、間接効用関数は t の関数で次の (関数) 方程式を満たさなければならない。

$$V(t) = \lambda \cdot \max[\delta^{-1} W(t), \beta EV(t+1) - c] + (1 - \lambda)[\beta EV(t+1) - c] \quad (3)$$

(3) 式の両辺の W について期待値をとると、モデルの定常性の仮定により $EV(t) = EV(t+1) \equiv V$ とおけば

$$V = \delta^{-1} \lambda E \max(W, R) + \delta^{-1} (1 - \lambda) R \quad (4)$$

となる。ここで $R = \delta(\beta V - c)$ である。したがって、雇用機会を受け入れるかの判断基準を、

$$W(t) > R \equiv \delta(\beta V - c) \quad (5)$$

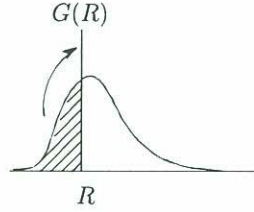


図 3: 留保賃金 R と $G(R) = P(W < R)$

と求めることができる。なお、

$$E \max(W, R) = \int_R^{\infty} W dG(W) + RG(R) \quad (6)$$

となることも記しておく。

2.2 期間構造を入れた求職モデルの推定：教育から就業への推移

Wolpin(1987)¹²の研究はもっとも簡単な求職モデルを一般化して、高校卒業生('79 NLS youth cohort)の求職活動における継続依存度(duration dependence)を実際に推定している。雇用機会を得る確率を t に依存させたモデルとなっているので λ_t と表記する。このとき $t = 1, \dots, T-1$ に対して間接効用関数は、

$$V(t) = \lambda_t \cdot \max[W_t, \beta EV(t+1) - c] + (1 - \lambda_t)[\beta EV(t+1) - c] \quad (7)$$

となる。しかし、ある期間 T 以降はオファーされた任意の賃金を受け入れると仮定し、期間 $(T, T + \tau)$ に就業する可能性の下で T 期の効用の計算は、

$$\begin{aligned} V(T) = & \lambda_T W_T - (1 - \lambda_T)c + \beta(1 - \lambda_T)[\lambda_{T+1}W_{T+1} - (1 - \lambda_{T+1})c] \\ & + \beta^2(1 - \lambda_T)(1 - \lambda_{T+1})[\lambda_{T+2}W_{T+2} - (1 - \lambda_{T+2})c] + \dots \\ & + \beta^\tau(1 - \lambda_T) \dots (1 - \lambda_{T+\tau-1})[\lambda_{T+\tau}W_{T+\tau} - (1 - \lambda_{T+\tau})c] \end{aligned} \quad (8)$$

より与えられる。したがって、 T 期から反復的(recursive)に $EV(t)$ を求めることができる。

推計上の議論では賃金が観測誤差を伴う状況(error in variable)を仮定し、

$$\bar{W}_t^i = W_t^i + \epsilon_t^i = x^{i'}\theta + u_t^i + \epsilon_t^i \quad (9)$$

¹²Wolpin, K.(1987), "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, Vol. 55, 801-817.

という計量モデルを考えている。観測誤差を入れたのは $w > R(\theta)$ と母数への制約がかかることを回避するためであり、観測誤差を仮定した方が都合が良いという（計量経済分析では希な）場合であると云えよう。期間依存度を示す λ_j の母数化 (parameterization) は j を j 回目の非雇用期間として

$$\lambda_j = \Phi(m_0 + m_1 j), \quad \Phi(\cdot) \text{ は標準正規分布関数} \quad (10)$$

と仮定する。例えば期間 $t = 2$ に雇用された人の尤度関数は、

$$L = [\lambda_1 P(W_1 < \beta EV(2) - c) + (1 - \lambda_1)] \quad (11) \\ \times \lambda_2 f(\bar{W}_2 | W_2 > \beta EV(3) - c) P(W_2 > \beta EV(3) - c)$$

となり、第1項は1期目にオファーがきたが断った、またはこなかったことに対応し、第2項(2期目)の $f(\cdot)$ はオファーを受け入れたことを所与とした \bar{W}_2 の密度である。Wolpin は (u_t, ϵ_t) が独立に正規分布に従うとして尤度を閉じた形で表現し、非雇用期間が長引くと雇用機会が減少する ($\hat{m}_1 < 0$) ことや、平均的な非雇用期間 (≈ 46 週) や留保賃金率 (≈ 80 ドル/卒業した週) を推計している。

2.3 選択売買権としての特許：ヨーロッパ型特許権行使株を保有する価値の推定

次に Pakes(1986)¹³ の求職モデルを応用した特許使用权の更新計画に関する研究を取りあげる(第二次大戦後の仏・英・独を対象に各種特許 (different cohorts of patents) の集計データを利用している)。この研究では期間 t の便益を表す式は、変数

- T 特許の有効期間 (Life of a patent) ,
- r_t 収益 (Return): 確率変数 W_t に相当 ,
- c_t 更新費用 (Renewal fee) ,
- r_{t+1} の分布は r_t, t , 母数 θ に依存する。

と置けば、

$$V(t) = \max[0, r_t + \beta EV(t+1) - c_t], \quad t = 1, 2, \dots, T-1 \quad (12)$$

と表現される。ここで \max 演算子の第1項は、 t 回目更新を止めた時には2度とその特許は使用できないことを、第2項は更新した際に特許からの現在収益と特許保有の期待現在価値が

¹³Pakes, A. (1986), "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, Vol. 54, 755-784.

得られること示している。法令により有効期間 T (16 ~ 20 年) は有限なので、

$$V(T) = \max[0, r_t - c_t] \quad (13)$$

となる。 $T = 3$ のとき具体的には、

$$\begin{aligned} V(1) &= \max[0, r_1 + \beta E[V(2)|r_1] - c_1], \\ V(2) &= \max[0, r_2 + \beta E[V(3)|r_2] - c_2], \\ V(3) &= \max[0, r_3 - c_3] \end{aligned} \quad (14)$$

となる。このモデルも定常性に欠けるので Wolpin のモデルと同様に、(14) は最後の式からまず $E[V(3)|r_2]$ を求め、recursive に解かねばならない。ここで、Pakes(1986) の重要な貢献のひとつは上式のように r_{t+1} と r_t が独立ではない下での考察であり、止め時 (optimal stopping) の基準を $r_t < \bar{r}_t$ (\bar{r}_t が唯一ある) で示している。したがって同時分布 $P(r_t \geq \bar{r}_t, \dots, r_2 \geq \bar{r}_2, r_1 \geq \bar{r}_1), t = 1, \dots, T$ を評価¹⁴ すれば各期の解約確率 $\pi(t, j)$ (j は特許の種類) が求まる。これより集計データ (各期の解約数 $n(t, j)$) で対数尤度 ($\log L = \sum_j \sum_t n(t, j) \log \pi(t, j)$) を組むことが出来て、現実の (各国各期の) 更新率を推計より再現することを試みている。

2.4 求職モデル (Search Model)-連続時間 (Continuous Time)

ここでは、離散時間の定常的求職モデルに対応して連続時間でのモデルを考えておこう。生存時間密度を用いる際は連続型モデルが有用であり、指数分布を用いた理論研究としては Lippman=McCall(1976)¹⁵ などが挙げられる。離散型モデルと同様に

- T 雇用機会を得るまでの時間 (time until wage offer arrives) $\sim \lambda e^{-\lambda t}$,
- G 賃金の分布関数 (distribution function of wage),
- c 求職コスト (search cost)
- $e^{-\delta t}$ t 時間後の割引現在価値 (present worth of 1 dollar obtained t later)

とすると、 t 時点の間接効用関数は、

$$V(t) = \max[\delta^{-1}W(t), K] \quad (15)$$

¹⁴ある種の密度 $f(r_{t+1}|r_t)$ 設定では $\{\bar{r}_t\}_{t=1}^T$ が構造母数 θ の関数となる。そして、収益流 r_t のデータを生成し $\{r_t \geq \bar{r}_t, \dots, r_2 \geq \bar{r}_2, r_1 \geq \bar{r}_1\}$ となる事象を数えて (Simulated frequency approach by Lerman=Manski(1981))、同時分布を評価する方法がとられている。

¹⁵Lippman=McCall(1976), "Job Search in a Dynamic Economy," *Journal of economic theory*, Vol. 12, 365-390.

ここで、

$$K = \int_0^{\infty} \left\{ \left[e^{-\delta s} EV(t+s) - c \int_0^s e^{-\delta t} dt \right] \lambda e^{-\lambda s} \right\} ds \quad (16)$$

である。したがって、期待効用は

$$V = \delta^{-1} E \max(W, R) \quad (17)$$

となる。さらに $V = EV(t) = EV(t+s)$ より、 $K = \lambda(\lambda + \delta)^{-1}V - (\lambda + \delta)^{-1}c$ となるので

$$R \equiv \delta K = -c + \delta^{-1} \lambda \int_R^{\infty} (W - R) dG(w) \quad (18)$$

が導出される。ここで $R + c$ は留保賃金 R を断った機会費用なので、上式は求職を続ける期待限界便益がその費用と均等する条件を示していると考えられる。この条件より R が求まり、非雇用期間が t 時間でその後就業した労働者の尤度関数は、母数 (λP) の指数分布を用いて (ハザード率が一定より)、

$$L = e^{-\lambda P t} \lambda P, \quad P = P(W > R) \quad (19)$$

により求めることができる。

2.5 GMC バスエンジンの最適交換モデル：Harold Zurcher の実証モデル

最後に構造的生存時間計量経済モデルとして著名な Rust(1987)¹⁶ の分析を取りあげてみよう。Rust は GMC 整備主任の Zurcher 氏よりデータの提供を受けて、同氏が動学的最適化行動をとっていたかの検証を試みている。この分析では問題は

- x_t 観測される状態変数 (Observable state variable) (↔ 求職モデルの W_t),
- ϵ_t 観測できない誤差 (Unobservable error)(↔ 雇用機会があれば 1、なければ 0),
- i_t 2 項決定関数 (Binary decision variable)(↔ 受容なら 1、そうでないなら 0),
- i_t^* 最適な i_t の値

により設定される。ここでより具体的には x_t はバスの走行マイル、またエンジンを分解修理 (overhaul) した時 $i_t = 1$ (その他ゼロ) である。負の便益関数 (よって低い値ほど良い) を、

$$V(x_t, \epsilon_t, i_t) = L(x_t, \epsilon_t, i_t) + \beta E[V^*(x_{t+1}, \epsilon_{t+1}) | x_t, \epsilon_t, i_t] \quad (20)$$

¹⁶Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, Vol. 55, 999-1033.

とおく。ここで状態評価関数 (value function) は、

$$V^*(x_{t+1}, \epsilon_{t+1}) = V[x_{t+1}, \epsilon_{t+1}, i_{t+1}^*(x_{t+1}, \epsilon_{t+1})]$$

と表している。この分析では攪乱項 ϵ_t に関して、

$$E[V^*(x_{t+1}, \epsilon_{t+1})|x_t, \epsilon_t, i_t] = E_x[E_{\epsilon}V^*(x_{t+1}, \epsilon_{t+1})|x_t, i_t] \equiv E_x[V^*(x_{t+1})|x_t, i_t] \quad (21)$$

を仮定している ($V^*(x_{t+1}) \equiv E_{\epsilon}V^*(x_{t+1}, \epsilon_{t+1})$ は単なる書き換えである)。この仮定は Rust モデルを解く上で重要であり、評価関数 $V^*(x_{t+1})$ が ϵ_t には依存しなくなる¹⁷。損失 L については、

$$L(x_t, \epsilon_t, i_t) = i_t[RC + c(0) + \epsilon_2(t)] + (1 - i_t)[c(x_t) + \epsilon_1(t)], \quad (22)$$

と定める。第1項は修理した際は交換費用 (RC) がかかるが走行マイルがゼロに復活することを示している ($\epsilon_{1,2}$ は Zurcher 氏が知るその他の状態)。費用 $c(x)$ は、

$$c(x) = \frac{\theta}{91 - x}, \quad x = 0, 1, \dots, 90 \quad (23)$$

などで与えられる (例えば走行マイルが延びると思わぬエンジン故障による費用がかかるであろう)。ここで走行マイルを90個に離散分割する工夫がなされていることに注意しよう (5000マイルごとに、 $x_t \in [0, 5,000) \Rightarrow x = 0, \dots, x_t \in [450,000, \infty) \Rightarrow x = 90$)。したがって遷移式も離散的に、

$$x_{t+1} = x_t + \begin{bmatrix} 0 \\ 1 \\ 2 \sim 90 \end{bmatrix}, \quad \text{をそれぞれ確率} \begin{bmatrix} p_0 \\ p_1 \\ 1 - p_0 - p_1 \end{bmatrix} \text{でとる。}$$

と表され、 (p_0, p_1) を母数として遷移確率 $p(x_{t+1}|x_t, i_t)$ が定められる。

さて Rust モデルの解を計算する。 $V^*(x)$ は x が $j = 0, 1, \dots, J (= 90 \text{ など})$ と変化するごとに V_j^* の値を取ると仮定する。したがって、条件付き期待値は容易に、

$$\begin{aligned} E[V^*(x_{t+1})|x_t, i_t] &= \sum_j V_j^* P(x_{t+1} = j|x_t, i_t) \\ &\equiv H(\{V_j^*\}, x_t, i_t) \end{aligned} \quad (24)$$

となる ($\{V_j^*\}$ はすべての V_j^*)。そして、

$$\text{Minimize}_{i_t} L(x_t, \epsilon_t, i_t) + \beta H(\{V_j^*\}, x_t, i_t) \quad (25)$$

¹⁷通常、条件付き期待値に関して $E\phi(X, Y) = E_Y[E\phi(X, Y|Y)]$ とするが条件付け変数から ϵ_t を除くため、条件付き独立 (\Rightarrow 独立は真ではない) を仮定している (遷移確率について $p(x_{t+1}, \epsilon_{t+1}|x_t, \epsilon_t, i) = q(\epsilon_{t+1}|x_{t+1})p(x_{t+1}|x_t, i)$ としている)。

すなわち上式を最小化する $i_t^*(x_t, \epsilon_t, \{V_j^*\})$ が最適な 2 項決定関数となる¹⁸。これを初出の (20) 式に代入して、

$$V^*(x_t, \epsilon_t, i_t^*) = L(x_t, \epsilon_t, i_t^*) + \beta H(\{V_j^*\}, x_t, i_t^*) \quad (26)$$

当期の評価関数を得る。 ϵ_t について期待値をとれば、

$$V^*(x_t) = E_\epsilon L(x_t, \epsilon_t, i_t^*) + \beta E_\epsilon H(\{V_j^*\}, x_t, i_t^*), \quad (27)$$

もしくは、 x_t も $k = 1, \dots, J$ 状態あるとして、

$$\begin{aligned} V_k^* &= E_\epsilon L[k, \epsilon_t, i_t^*(k, \epsilon_t, \{V_j^*\})] \\ &\quad + \beta E_\epsilon H[\{V_j^*\}, k, i_t^*(k, \epsilon, \{V_j^*\})] \end{aligned} \quad (28)$$

となり、 J 個の連立方程式を解けば解が求まる。

2.5.1 簡略化した Rust モデル

Rust モデルは既に見てきたように一般にはかなり複雑で、明示的に 2 項決定関数 $i_t^*(x_t, \epsilon_t, \{V_j^*\})$ を解くことができない。そこで、 i_t^* の定性分析ができることとなる簡略化したモデルを用いて彼が求めた最適解の性質を調べてみよう。その為にここでは

$$\begin{aligned} &\text{損失関数 (費用 } c \text{ 一定)} \quad L(x_t, \epsilon_t, i_t) = (c + \epsilon_t)i_t + x_t, \\ &x \text{ は 3 つの値をとる } (x = 0, 1, 2), \\ &x_t = 2 \Rightarrow i_t = 1 \text{ (上限に達すると必ず修理)}, \\ &\text{遷移式 } i_t = 0 \Rightarrow x_{t+1} = x_t + w, \quad i_t = 1 \Rightarrow x_{t+1} = w, \\ &1/2 \text{ の確率で } w=1, \text{ 同じく } 1/2 \text{ の確率で } w=0 \end{aligned}$$

と変数を設定する。このとき負の便益関数と来期の評価関数は、

$$V(x_t, \epsilon_t, i_t) = (c + \epsilon_t)i_t + x_t + \beta E_x[V^*(x_{t+1})|x_t, i_t], \quad (29)$$

$$V^*(x_{t+1}) = i_t V^*(w) + (1 - i_t) V^*(x_t + w) \quad (30)$$

¹⁸Rust は $\{i_t^*, x_t\}$ を観測するので、 $(\epsilon_1(t), \epsilon_2(t))$ が独立に TypeI の極値分布に従うとすれば 2 項決定確率は $P(i = 1|x, \theta) = \exp(L(x, 1, \theta) + \beta E_\theta V(x, 1)) [\exp(L(x, 1, \theta) + \beta E_\theta V(x, 1)) + \exp(L(x, 0, \theta) + \beta E_\theta V(x, 0))]^{-1}$ (ロジットモデル) で与えられる。これより尤度は $\Pi_t P(i_t|x_t, \theta) p(x_t|x_{t-1}, i_{t-1})$ となるが、尤度を評価する毎に $E_\theta V$ を収束させる計算法 (Nested fixed-point ML) で推定量 $\hat{\theta}$ を得ている。

である。 (x_t, i_t) を所与として w について期待値をとると、

$$E[V^*(x_{t+1})|x_t, i_t] = i_t \left[\frac{1}{2}V^*(0) + \frac{1}{2}V^*(1) \right] + (1 - i_t) \left[\frac{1}{2}V^*(x_t) + \frac{1}{2}V^*(x_t + 1) \right] \quad (31)$$

以上3つの式から、

$$\begin{aligned} V(x, \epsilon, 1) &= c + \epsilon + x + \beta \left[\frac{1}{2}V^*(0) + \frac{1}{2}V^*(1) \right], \\ V(x, \epsilon, 0) &= x + \beta \left[\frac{1}{2}V^*(x) + \frac{1}{2}V^*(x + 1) \right] \end{aligned}$$

となる。したがって $x_t \neq 2$ を仮定すると、もし、

$$c + \epsilon + \frac{\beta}{2}[V^*(0) + V^*(1)] < \frac{\beta}{2}[V^*(x) + V^*(x + 1)]$$

ならば $i^*(x_t, \epsilon_t) = 1$ (V の値は低い方がよい)。また、 $c + \epsilon > 0$ を仮定すると、もし $x_t = 0$ ならば $i_t^* = 0$ 。 $x_t = 1$ のとき、もし $c + \epsilon_t < \frac{\beta}{2}[V^*(2) - V^*(0)]$ ならば $i_t^* = 1$ である (すなわち、 $x_t \neq 2, c + \epsilon > 0$ の下では、 $x_t = 1$ のときに $i_t^* = 1$ (修理) の可能性がある)。

そこで、 $V^*(2) - V^*(0)$ を評価しよう。(28) 式に (30) 式、 $i_t = i_t^*$ を代入すると

$$\begin{aligned} V(x_t, \epsilon_t, i_t^*) &= (c + \epsilon_t)i_t^* + x_t + \frac{\beta}{2}[V^*(0) + V^*(1) - V^*(x_t) - V^*(x_{t+1})]i_t^* \\ &\quad + \frac{\beta}{2}[V^*(x_t) + V^*(x_{t+1})] \end{aligned} \quad (32)$$

となる。ここで $V^*(x_t) = E_\epsilon V(x_t, \epsilon_t, i_t^*)$ に注意すると、 $x_t = 0$ ときには $i_t^* = 0$ なので、

$$V^*(0) = \frac{\beta}{2}[V^*(0) + V^*(1)] \quad (33)$$

を得る。以上の設定より $x_t = 2$ ならば $i_t^* = 1$ 、さらに $E\epsilon_t = 0$ を仮定すれば、

$$V^*(2) = c + 2 + \frac{\beta}{2}[V^*(0) + V^*(1)] \quad (34)$$

である。よって上の2つの式から $V^*(2) - V^*(0) = c + 2$ と評価できる。これを先の条件に代入すると以下の結果を得る。

$$x_t = 1 \text{ のとき、もし } \epsilon_t < \left(\frac{\beta}{2} - 1\right)c + \beta \text{ ならば } i_t^* = 1 \quad (35)$$

これより定性分析として $x_t = 1$ のとき $((\beta/2 - 1) < 0$ に注意し不等式が満足されるかを考えると)

より大きな c は、 $i_t^* = 0$ を選びやすい、
より大きな β は、 $i_t^* = 1$ (修理) を選びやすい

という結論が得られる。

2.5.2 コメント

Rust の論文では以上で議論したモデルの最適解により実際のデータを分析している。実証分析の結論は現実のデータは最適解と整合的であるとするものであるが、その評価は計量経済理論の分析に比べると議論の余地があろう。

第5回：左打ち切り (Left Censoring) 問題

2005年6月1日

概要

デュレーション (duration) モデルでは観測者が一部の経過期間 (spell) を途中から計測し始める場合に左打ち切り (left censoring) 問題が発生する。その対処として経過期間の開始時点分布 (starting time distribution) を導入する。Amemiya(1999)¹ は左打ち切り問題の分類や、開始時点分布の扱いに関係するいくつかの最尤法 (最尤法/条件付最尤法/セミパラメトリック最尤法) に対して効率性の比較を行っているので、当論文に沿う形で左打ち切り問題を解説する。

1 はじめに

まず左打ち切り (left censoring) を理解するため生存時間解析 (survival analysis) の例として著名なスタンフォード大学病院での心臓移植を例として考えよう。ここで関心のある寿命変数 $T \geq 0$ は術後の生存期間で、密度関数と分布関数をそれぞれ $f(t)$, $F(t)$ とする。ここで当大学病院では1970年頃より心臓移植が始められたので、生存期間の観測を始めた日付 (calendar time) を1970年として、これを a にする。この場合には初めて手術が行われた日付はデータよりはっきり確定できるとも考えられる。ところが、例えば失業期間の分析などでは観測を開始した時点ではいつから失業が発生したか不明なことがしばしば起きる。すなわち観測開始時点 (これを0とおく) において既に以前より寿命分布に関わる事象が発生していると考えられることが一般的である。こうした場合は左打ち切り (left censoring) が発生していると云うが、打ち切りを考慮せずに寿命分

¹Amemiya, T. (1999), "A Note on Left Censoring," in C. Hsiao, K. Lahiri, L-F. Lee, and M. Pesaran, eds., *Analysis of Panels and Limited Dependent Variable Models*, Cambridge University Press.

布に関する母数の推定を行うと推定量は一般に一致性 (consistency) を持たなくなる。したがって観測時点で寿命事象の開始時点 follow up する必要が生じる。

失業、あるいは手術などの事象の開始時点 (starting time) を確率変数 X とし、その密度 $h(x)$ の台 (support) は観測開始の時刻 0 を含む区間 $[a, b]$ としよう。ここで仮定として生存時間 T と開始時点 X は独立であり、各個体の標本 (x_i, t_i) に対しては同質的母集団 (homogeneous population) を想定、すなわち $h(x)f(t)$ からの i.i.d. 標本としよう。

この設定では、まず時刻 a (例えば 1970 年) 以前には心臓移植は行われていなかったと考え、また観測を始めた時刻 0 (例えば 1980 年) 以前に移植手術を受けた患者 i に対しては、手術の開始時点 x_i や生存期間 t_i が観測されないと仮定しよう。図 1 では観測期間を $[0, b]$ (例えば $b = 2004$ 年) とした場合、代表的な 3 種類の経過期間を用いて、左打ち切りの分類を示してある。

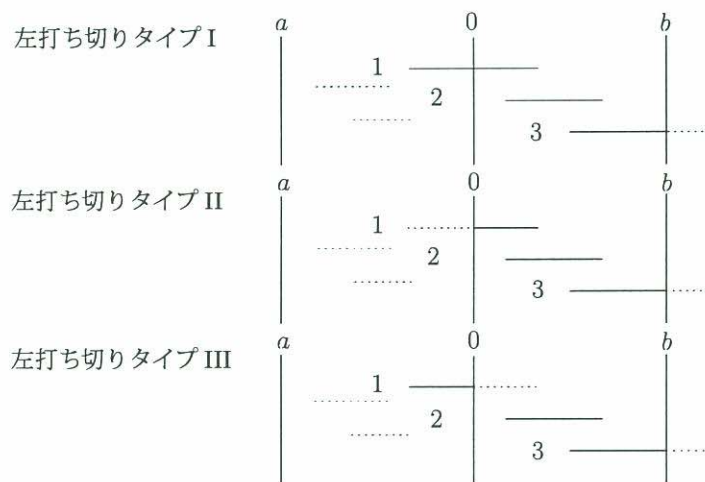


図 1: 3 タイプの左打ち切り (点線は観測されない)

各「左打ち切りタイプ」において経過期間の第 2 種は、観測期間中に開始時点 x_i と生存期間 t_i がともに観測された完全な経過期間 (complete spell) である (観測期間中には関心のある事象 (死亡) が確認でき、転院などは起きないとする)。第 3 種は、観測終了による右打ち切り (right censoring) を表している。

注意点として、右打ち切りとの対応から「打ち切りタイプ II」のみが

(しばしば統計学においても) 打ち切り問題とされるが、本来3タイプに分類する必要がある。心臓移植の例では、「タイプII」の左打ち切りはいつ手術が施されたわからない状況を(しかし、これは考えにくい)、「タイプIII」では時刻0における右打ち切り(例えば、他の臨床試験により観察対象から外された)を表している。「タイプI」では、完全な経過期間が観察されている。

左打ち切りが問題となるのは、いずれのタイプにおいても短い経過期間(点線)は時刻0に到達できず、長い経過期間ほどより観察されやすい(selectivity bias)という場合であり、平均生存時間は過大評価されるであろう。そこで、先に導入した開始時点 X を用いて各経過期間が観察される確率を評価する。第1種とその他経過期間が観察される事象 $A_1 = \{x, t \mid t > -x, 0 > x > a\}$, $A_2 = \{x, t \mid x > 0\}$ は排反で、

$$\begin{aligned} \text{経過期間 1: } P_1 \equiv P(A_1) &= \int_a^0 h(x)[1 - F(-x)]dx, \\ \text{経過期間 2, 3: } P_2 \equiv P(A_2) &= \int_0^b h(x)dx \end{aligned}$$

となる。したがって、ある経過期間が観察される確率は $P \equiv P_1 + P_2$ である(他方、観察されない場合は $x_i + t_i > 0$ のとき)。 N 人中 n 人の経過期間が観察されたとき、各「打ち切りタイプ」に応じた正しい尤度関数(full LF)は、

$$\begin{aligned} L_1 &= \prod_1 h(x_i)f(t_i) \prod_2 h(x_i)f(t_i) \prod_3 h(x_i)[1 - F(b - x_i)] \prod_{all} P^{-1}, \\ L_2 &= \prod_1 \int_a^0 h(x)f(t_i - x)dx \prod_2 h(x_i)f(t_i) \prod_3 h(x_i)[1 - F(b - x_i)] \prod_{all} P^{-1}, \\ L_3 &= \prod_1 h(x_i)[1 - F(t_i)] \prod_2 h(x_i)f(t_i) \prod_3 h(x_i)[1 - F(b - x_i)] \prod_{all} P^{-1} \end{aligned}$$

と考えられる。各尤度において第2項と第3項は、完全な経過期間と右打ち切りに対応して共通し、左打ち切りに対応した第1項のみが異なる(L_1 においては、第1項と第2項が区別されていないことに注意する)。ここで重要なことは、各尤度ともすべて第4項 $P_1 + P_2 = P(i = 1, \dots, n)$ により条件付けられている(割られている)ことである。「打ち切りタイプI」

の尤度 L_1 を書き直すと、

$$L_1 = \prod_1 h(x_i)f(t_i)/P_1 \prod_2 h(x_i)f(t_i)/P_2 \prod_3 h(x_i)[1 - F(b - x_i)]/P_2 \\ \times \prod_1 P_1/(P_1 + P_2) \prod_2 P_2/(P_1 + P_2) \prod_3 P_2/(P_1 + P_2) = L_{11} \times L_{12}$$

となる。一見、 L_{11} のように各経過期間 1,2,3 が観察される確率で割った方が良さそうであるが、この際は L_{12} を無視 (各経過期間の第何種で観察されたのかを) した部分尤度 (partial LF) となるので、一般に一致性は有しても効率性に劣る (厳密な議論は、Amemiya(1999) を参照)。

2 最尤推定量

以下に続く各節では、「左打ち切りタイプ I」について最尤推定量の漸近的性質を見ていく。簡単化のため、観測される経過期間は左打ち切りされた第 1 種に限定しよう。したがってある経過期間が観測される事象を $A \equiv A_1$ と置き直す。つまり、

$$A = \{x, t \mid t > -x, 0 > x > a\} \quad (1)$$

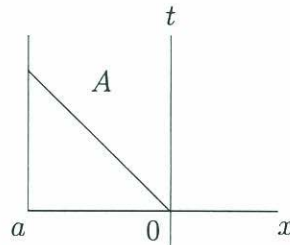


図 2: (t, x) 上での領域 A

である。したがって、経過期間が観察される確率は $P \equiv P(A) = \int_A h(x)f(t)dxdt = \int_a^0 h(x)[1 - F(-x)]dx$ で評価することとなる。加えて、ここでは関心のある母数 θ は 1 つ (scalar) で f のみに依存すると仮定しよう。このとき、観測数を n とした尤度は、

$$L = \prod_{i=1}^n h(x_i)f(t_i)P^{-1} \quad (2)$$

となる。不均一母集団 (heterogeneous population) や母数ベクトルへの拡張は、例えば、

$$L = \prod_{i=1}^n h(x_i - z_i' \beta) f(t_i - s_i' \theta) P_i^{-1} \quad (2^*)$$

とすることが可能である。ここで、 z_i, s_i は個体の属性を示す説明変数ベクトルで、 β, θ はそれに対応した母数ベクトル、 $P_i = \int_a^0 h(x - z_i' \beta) [1 - F(-x - s_i' \theta)] dx$ である。一般的なモデルへの拡張は $(h_i, f_i, P_i, a_i, b_i)$ と添え字 i を加えるのみで尤度の構成は同一となる。

ここで簡略化された左打ち切りモデル (2) 式に自然対数を取り (θ に依存しない h を無視して)、1階の微分をとると、

$$\log L = \sum_{i=1}^n \log f(t_i) - n \log P \quad (3)$$

$$\begin{aligned} \frac{\partial \log L}{\partial \theta} &= \sum_{i=1}^n \frac{1}{f} \frac{\partial f}{\partial \theta} - \frac{n}{P} \frac{\partial P}{\partial \theta} \\ &= \sum_{i=1}^N D_i \frac{1}{f} \frac{\partial f}{\partial \theta} - \sum_{i=1}^N D_i \frac{1}{P} \frac{\partial P}{\partial \theta} \end{aligned} \quad (4)$$

となる。ただし、 N は観測されなかった経過期間を含んだ標本総数で、 D_i は $(x_i, t_i) \in A$ (観測された) ならば 1 (その他でゼロ) をとるダミー変数である。

一般に最尤推定量は一致有効推定量なので、続く導出は必ずしも必要でないがその他の最尤推定量の比較のため具体的に求める。一致性については²、 $ED(1/f)(\partial f/\partial \theta) = \int_A (1/f)(\partial f/\partial \theta) h f dt dx$, $ED = P$ より、

$$E \frac{1}{N} \frac{\partial \log L}{\partial \theta} = \int_A \frac{1}{f} \frac{\partial f}{\partial \theta} h f dt dx - \frac{\partial P}{\partial \theta} = \frac{\partial}{\partial \theta} \int_A h f dt dx - \frac{\partial P}{\partial \theta} = 0 \quad (5)$$

である。さらに (4) 式を θ についてもう一度微分して、

$$\frac{\partial^2 \log L}{\partial \theta^2} = - \sum_{i=1}^n \frac{1}{f^2} \left[\frac{\partial f}{\partial \theta} \right]^2 + \sum_{i=1}^n \frac{1}{f} \frac{\partial^2 f}{\partial \theta^2} + \frac{n}{P^2} \left[\frac{\partial P}{\partial \theta} \right]^2 - \frac{n}{P} \frac{\partial^2 P}{\partial \theta^2} \quad (6)$$

²Amemiya(1985), *Advanced Econometrics*, Blackwell. p.110 の定理 4.1.2 を参照。

が得られる。上式の右辺の期待値をとると第2項 $\frac{\partial^2}{\partial \theta^2} \int_A f h dt dx$ と第4項 $-\partial^2 P / \partial \theta^2$ が相殺されて、

$$-E \frac{1}{N} \frac{\partial^2 \log L}{\partial \theta^2} = E \left[\frac{\partial \log f}{\partial \theta} \right]^2 - \frac{1}{P} \left[\frac{\partial P}{\partial \theta} \right]^2 \quad (7)$$

となる。一般には $\sqrt{N}(\hat{\theta} - \theta)$ の漸近分散は (7) 式の逆数で与えられる³。

3 条件付き最尤推定量

次に初期分布 $h(x)$ の情報に依存せず θ の一致推定量を求める手法を考えよう。 x を所与した条件付き密度 $f(t|x)$ に基づく最尤法は、Lancaster(1979)⁴ により失業期間の分析で用いられている。この手法が有用であるのは、研究者にとって多くの場合、適切に $h(x)$ を特定化することは困難を伴うからである。

左打ち切り問題の困難な点は、観察される確率 P で正しく条件付けて (割って) 一致性を示すこと以上に、 $h(x)$ を特定化しなければならないことにある。仮に h が θ を依存していない場合でも、 h を無視した尤度は $L = \Pi f / P(h)$, (ここで $P(h) = \int_A h(x) f(t) dx dt$) となるので h はやはり規定しなければならない。先の心臓移植の例では、 $[b, 0]$ 上 (1970年から1980年) に x が一様に分布している (constant entry rate) と考えても良いかもしれない。しかし経済学において失業期間の分析や、銀行の倒産を対象とする際に (創業の) 密度 h の特定化 (これは a を定めることも意味する⁵) は難しいであろう。

Lancaster(1979) の手法においても、引き続き同質的母集団を仮定の下で開始時点 x_i が与えられたときの条件付き密度は $A_i \equiv \{t > -x_i\}$ を用いて、

$$f(t_i|x_i) = \frac{h(x_i) f(t_i)}{\int_{A_i} h(x_i) f(t) dt} = \frac{f(t_i)}{1 - F(-x_i)}$$

³Amemiya(1985),p.121 を参照。

⁴Lancaster,T.(1979), "Econometric Methods for the Duration of Unemployment," *Econometrica*, Vol. 47, 939-956.

⁵ a は関心のある経過期間の歴史の始まりなので、本邦の近代的銀行の創業は明治維新 (1868年) 以降であろう。ちなみに世界的に見れば銀行業は古代アテナイの時代 (紀元前5世紀) にも存在していた。

となるので h に依存しないことがわかる。条件付き尤度は

$$L^* = \prod_{i=1}^n \frac{f(t_i)}{1 - F(-x_i)} \quad (8)$$

不均一母集団の場合は例えば、

$$L^* = \prod_{i=1}^n \frac{f(t_i - s'_i \theta)}{1 - F(-x_i - s'_i \theta)} \quad (8^*)$$

と表せる。次式 (9) の条件付き対数尤度に対し θ について 1 階の微分をとれば

$$\log L^* = \sum_{i=1}^n \log f(t_i) - \sum_{i=1}^n \log[1 - F(-x_i)] \quad (9)$$

$$\frac{\partial \log L^*}{\partial \theta} = \sum_{i=1}^n \frac{1}{f} \frac{\partial f}{\partial \theta} + \sum_{i=1}^n \frac{1}{1 - F} \frac{\partial F}{\partial \theta} \quad (10)$$

となる。推定量の一致性については、前節の議論と同様に標本総数 N になることに注意して、

$$\begin{aligned} E \frac{1}{N} \frac{\partial \log L^*}{\partial \theta} &= \int_A \frac{1}{f} \frac{\partial f}{\partial \theta} h f dt dx + \int_A \frac{1}{1 - F} \frac{\partial F}{\partial \theta} h f dt dx \quad (11) \\ &= \frac{\partial}{\partial \theta} \int_a^0 h(x) [1 - F(-x)] dx + \frac{\partial}{\partial \theta} \int_a^0 h(x) F(-x) dx \\ &= \frac{\partial}{\partial \theta} \int_a^0 h(x) dx = 0 \end{aligned}$$

となることより示すことができる。ただし、2 番目の等式では $\int_{-x}^{\infty} f dt = 1 - F(-x)$ を用いた。次に (10) 式を再度 θ について微分すると

$$\begin{aligned} \frac{\partial^2 \log L^*}{\partial \theta^2} &= - \sum_{i=1}^n \frac{1}{f^2} \left[\frac{\partial f}{\partial \theta} \right]^2 + \sum_{i=1}^n \frac{1}{f} \frac{\partial^2 f}{\partial \theta^2} + \sum_{i=1}^n \frac{1}{(1 - F)^2} \left[\frac{\partial F}{\partial \theta} \right]^2 \quad (12) \\ &\quad + \sum_{i=1}^n \frac{1}{1 - F} \frac{\partial^2 F}{\partial \theta^2} \end{aligned}$$

となる。ここで第 2 項が $-\frac{\partial^2}{\partial \theta^2} \int_0^a h f dx$ となるので (第 4 項) = -(第 2 項) が相殺されて

$$-E \frac{1}{N} \frac{\partial^2 \log L^*}{\partial \theta^2} = \int_A \frac{1}{f} \left[\frac{\partial f}{\partial \theta} \right]^2 h dt dx - \int_a^0 \frac{1}{1-F} \left[\frac{\partial F}{\partial \theta} \right]^2 h dx \quad (13)$$

となる。

ところで、最尤推定量 $\hat{\theta}$ と条件付き最尤推定量 $\hat{\theta}_c$ の漸近分散の比較は、(7) 式と上式 (13) の大小を調べればよいが、これは以下のコーシー・シュワルツの不等式 (Cauchy-Schwartz inequality) により示される。

$$E_x U^2 E_x V^2 \geq (E_x UV)^2 \quad (14)$$

ここで、 $E_x(Z) = \int_a^0 z h(x) dx$ と定め、 $U = \frac{1}{\sqrt{1-F}} \frac{\partial F}{\partial \theta}$ 、 $V = \sqrt{1-F}$ ととれば、(7) 式と (13) 式の第 2 項の関係 (第 1 項は共通) は $\frac{1}{F} \left[\frac{\partial F}{\partial \theta} \right]^2 \leq \int_a^0 \frac{1}{1-F} \left[\frac{\partial F}{\partial \theta} \right]^2 h dx$ となり、(7) 式が大きい。したがって、条件付き最尤推定量 $\hat{\theta}_c$ は一般に有効とはならない。

4 $h(x)$ と θ のセミパラメトリック推計

Goto (1993,1996)⁶ は、(2) 式に相当するモデルで密度 $h(x)$ と θ (f はこの母数に依存する) に対するセミパラメトリック最尤法を考察し、最大化の解 $h(x)$ を (2) 式に代入した際に条件付き尤度 (8) 式が得られることを証明している。したがって、Lancaster(1979) による条件付き最尤推定量 $\hat{\theta}_c$ は、(2) 式から求まるセミパラメトリック最尤推定量 $\hat{\theta}_s$ と同等であったという意味で重要な結果と考えられる。ここでその議論を説明しておこう。

まず (2) 式より

$$L = \prod_{i=1}^n \frac{h(x_i) f(t_i)}{\int_a^0 h(x) [1 - F(-x)] dx} \quad (15)$$

において、 $h(\cdot)$ に関して最大化する。分母は小さいほどよいから、 $h(\cdot)$ の取り方は図 3 に示す通り、観察された x_i 周辺の区間 d 以外ではゼロとすることができる。また、 $[a, 0]$ 上で $(\sum h_k d) = 1$ より、 $d = (\sum h_k)^{-1}$ ととる。

⁶Goto, F.(1993), "Consistency and Efficiency of Semiparametric Estimators in Left Censored Duration Models," *Unpublished Ph.D.Thesis*. Stanford University. Goto, F.(1996), "Achieving Semiparametric Efficiency Bounds in Left Censored Duration Models," *Econometrica*, Vol. 64, 439-442.

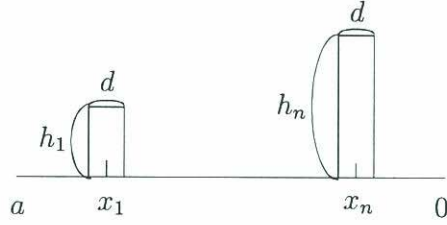


図 3: 階段関数 $h(\cdot)$ の取り方

そして、分母の積分値を以下のように近似する。

$$\int_a^0 h(x)[1 - F(-x)]dx \simeq \sum_k dh_k[1 - F(-x_k)] \quad (16)$$

このとき (15) 式は、

$$L = \prod_{i=1}^n \frac{h_i f(t_i)}{\sum_{k=1}^n h_k (\sum_j h_j)^{-1} [1 - F(-x_k)]} \quad (17)$$

と書き表される。 $h_k \equiv h_k / \sum h_k$ と置き直せば、次式の条件付最大化問題

$$L = \prod_{i=1}^n \frac{h_i f(t_i)}{\sum_{k=1}^n h_k [1 - F(-x_k)]} \quad \text{制約条件} \quad \sum_k h_k = 1 \quad (18)$$

に置き換わる (d は定数とみなす)。まず制約を考慮せず上式に対数をとると、 h_i に関する項は、

$$\sum \log h_i - n \log \sum h_k a_k, \quad \text{ただし} \quad a_k = 1 - F(-x_k) \quad (19)$$

各 h_i に関する 1 階条件より、

$$\frac{1}{h_i} - \frac{na_i}{\sum h_k a_k} = 0 \quad (20)$$

$$h_i = \left(\frac{a_1}{a_i} \right) h_1, \quad i = 1, 2, \dots, n \quad (21)$$

である。ここで両辺の和より制約 $\sum_k h_k = 1$ を用いると次の解を得る。

$$h_i = \frac{1}{a_i \sum_k (a_k)^{-1}} \quad (22)$$

この解を (18) 式に代入・整理すれば

$$\frac{1}{n^n} \prod_{i=1}^n \frac{f(t_i)}{1 - F(-x_i)} \quad (23)$$

が得られる。上式は、 θ に依存しない部分を除いては条件付き尤度 (8) 式に他ならない。また、以上の議論は不均一母集団の場合でも成り立つ、つまり (2*) 式に相当するモデルでは h と β について最大化することで (8*) 式を得るであろう。

5 $h(x)$ の別個の推定

ここでは、 θ の推計に用いた n 個の標本とは独立な K 個の標本を利用して $h(x)$ を推定する場合を考える。例えば銀行倒産までの期間分析で $n=100$ 行の標本で θ を推計して、別途に $K=100$ 行の設立年次の標本を得ている状況である。 h を規定せずに独立な K 個の標本で開始時点の密度 $h(x)$ またはその分布 $H(x)$ を推計することで、条件つき最尤法よりも効率的な推定量 $\tilde{\theta}$ を求められるのではないかと考えられるのがこの状況設定の動機である。Nickell(1979)⁷ は同様な手法を用いたが、分離して $H(x)$ を推計することによる $\tilde{\theta}$ の漸近分布への影響は分析されていない。

そこでまず我々は、以下の W を最大化することを考えみよう。

$$W = \prod f(t_i) \hat{P}^{-1} \quad (24)$$

ここで、 $\hat{P} = \int_a^0 [1 - F(-x)] d\hat{H}$ 、 \hat{H} は K 個の標本に基づく開始時点 x の経験分布関数である (図 4 で $x_{(k)}$ は K 個の標本を小さい値順に並べたもの)。

したがって、標本上では、

$$\hat{P} = K^{-1} \sum_{i=1}^K \bar{F}(-x_i), \quad \text{ただし } \bar{F}(-x_i) = 1 - F(-x_i) \quad (25)$$

ここでの推定量 $\tilde{\theta}$ の漸近分布は

$$\sqrt{N}(\tilde{\theta} - \theta) = - \left(\frac{1}{\sqrt{N}} \frac{\partial \log W}{\partial \theta} \right) \left(\frac{1}{N} \frac{\partial^2 \log W}{\partial \theta^2} \right)^{-1} \quad (26)$$

⁷Nickell, S. (1979), "Estimating the Probability of Leaving Unemployment," *Econometrica*, Vol. 47, 1249-1266.

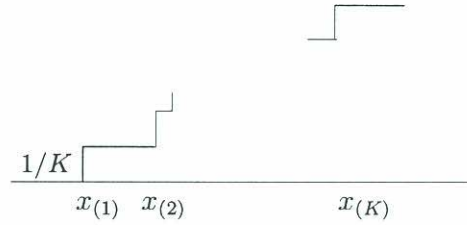


図 4: 経験分布関数 \hat{H} の例

より得ることが出来る⁸。

標本総数 N で除された 2 階微分の項は、 H を推計しない場合と同じ極限に収束することになるので 1 階微分の項についてのみ考えると、

$$\log W = \sum_{i=1}^n \log f(t_i) - n \log \sum_{i=1}^K \bar{F}(-x_i) \quad (27)$$

$$\begin{aligned} \frac{\partial \log W}{\partial \theta} &= \sum_{i=1}^n \frac{1}{f} \frac{\partial f}{\partial \theta} - \frac{n}{\sum_{i=1}^K \bar{F}} \sum_{i=1}^K \frac{\partial \bar{F}}{\partial \theta} \\ &= \sum_{i=1}^N \left(D \frac{1}{f} \frac{\partial f}{\partial \theta} - \frac{\partial P}{\partial \theta} \right) + N \left(\frac{\partial P}{\partial \theta} - K^{-1} \sum_{i=1}^K \frac{\partial \bar{F}}{\partial \theta} \right) \\ &\quad + \frac{N}{K} \sum_{i=1}^K \frac{\partial \bar{F}}{\partial \theta} \left(1 - \frac{n}{N} \frac{1}{K^{-1} \sum_{i=1}^K \bar{F}} \right) \end{aligned} \quad (28)$$

となる。ここで、 D は先で用いた経過期間 i が観察されたとき 1 をとるダミー変数である。上式 (28) では $\partial P / \partial \theta$ と $K^{-1} \sum_{i=1}^K \partial \bar{F} / \partial \theta$ に対し、引いて足し込む操作をした。また、上式 (28) の第 2 項に関して、

$$\begin{aligned} \frac{\partial P}{\partial \theta} - \frac{1}{K} \sum_{i=1}^K \frac{\partial \bar{F}}{\partial \theta} &= \frac{\partial}{\partial \theta} \left(P - K^{-1} \sum_{i=1}^K \bar{F} \right) = \frac{\partial}{\partial \theta} \int_a^0 [1 - F(-x)] d(H - \hat{H}) \\ &= \int_a^0 \frac{\partial F}{\partial \theta} d(\hat{H} - H) = \frac{1}{K} \sum_{i=1}^K \frac{\partial F}{\partial \theta} - E \frac{\partial F}{\partial \theta} \end{aligned} \quad (29)$$

⁸Amemiya(1985),p.111 を参照。

となるが、さらに (28) 式において $N/K \sum_{i=1}^K \bar{F}$ と $N/K \sum_{i=1}^K \partial \bar{F} / \partial \theta$ に関して線形で表されるように以下の操作を行う。

$$\begin{aligned} 1 - \frac{n}{N} \frac{1}{K^{-1} \sum_{i=1}^K \bar{F}} &= \frac{1}{K^{-1} \sum_{i=1}^K \bar{F}} \left(K^{-1} \sum_{i=1}^K \bar{F} - \frac{n}{N} \right) \\ &= \frac{1}{K^{-1} \sum_{i=1}^K \bar{F}} \left[\left(K^{-1} \sum_{i=1}^K \bar{F} - P \right) - \left(\frac{n}{N} - P \right) \right] \end{aligned} \quad (30)$$

よって、 $1/\sqrt{N}$ 倍した (28) 式は結局次式で表すことができる。

$$\begin{aligned} \frac{1}{\sqrt{N}} \frac{\partial W}{\partial \theta} &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(D \frac{1}{f} \frac{\partial f}{\partial \theta} - \frac{\partial P}{\partial \theta} \right) + \sqrt{N} \left(\frac{1}{K} \sum_i \frac{\partial F}{\partial \theta} - E \frac{\partial F}{\partial \theta} \right) \\ &+ K^{-1} \sum_{i=1}^K \frac{\partial \bar{F}}{\partial \theta} \frac{1}{K^{-1} \sum_{i=1}^K \bar{F}} \left[\sqrt{N} \left(K^{-1} \sum_{i=1}^K \bar{F} - P \right) - \sqrt{N} \left(\frac{n}{N} - P \right) \right] \end{aligned} \quad (31)$$

ここで、第2項と第3項は H を推計したために現れた項であるが、第1項と第4項 $\frac{\partial P}{\partial \theta} \frac{1}{P} [-\sqrt{N}(\frac{n}{N} - P)] (K \rightarrow \infty \text{とした})$ の和は、 $1/\sqrt{N}$ 倍した (4) 式の右辺と漸近的に等しい。したがって、(7) 式を行列 B (母数 θ をベクトルとした際に) で表しておく、独立性⁹ と $\bar{F} = 1 - F$ より $\sqrt{N}(\bar{\theta} - \theta)$ の漸近共分散行列は以下で与えられる。

$$B^{-1} \left[B + \frac{N}{K} V \left(\frac{\partial F}{\partial \theta} \right) + \frac{N}{K} \left(\frac{\partial \log P}{\partial \theta} \right)^2 V(F) - 2 \frac{N}{K} \left(\frac{\partial \log P}{\partial \theta} \right) \text{Cov} \left(F, \frac{\partial F}{\partial \theta} \right) \right] B^{-1} \quad (32)$$

すなわち、 $N/K \rightarrow 0$ のとき¹⁰、この推定量 $\tilde{\theta}$ は H を既知とした際の推定量 $\hat{\theta}$ と同等の効率性を有する。

不均一標本を扱う場合には、次式を最大化する。

$$W = \prod_{i=1} f(t_i - s'_i \theta) \hat{P}_i^{-1} \quad (24^*)$$

ここで、 $\hat{P}_i = \int_a^0 [1 - F(-x - s'_i \theta)] d\hat{H}(x - z'_i \beta)$ 、 \hat{H} は $\hat{u}_k = x_k - z'_k \beta$ において増分 $1/K$ をもつ階段関数で $\hat{\beta}$ は最小2乗推定量を用いればよい。し

⁹ θ の推計に用いる標本 n も用いて H を推計してもよいが、(第1項+第4項) と第2,3項は独立でなくなるので (32) 式で共分散項が増えるであろう。

¹⁰ 例えば $K = 100$ をとり $N = 10$ をとれると想定して、 $K = N^2$ か $K = 10 \times N$ とするかは見方による。また、 $K/N \rightarrow \text{定数}$ ならば共分散の大小は一概には云えない。 $K/N \rightarrow \infty$ のときは (26) 式の安定化の \sqrt{N} を変える必要があるだろう。

たがって、

$$\hat{P}_i = \frac{1}{K} \sum_{i=1}^K [1 - F(-\hat{u}_k - s'_i \theta - z'_i \beta)] \quad (25^*)$$

上式では、残差 \hat{u}_k を求めるための β の推定でさらに標本誤差が含まれることとなるが、収束の早さは同質的母集団の場合と変わらない。

他方、カーネル推定量で密度 h から推計しても、同じ漸近的な結果を得ることが出来る。カーネル推定量は

$$\hat{h}(x) = \sum_{i=1}^K g\left(\frac{x_i - x}{d}\right) \frac{1}{Kd} \quad (33)$$

で与えられるが、ここで、 g はカーネル関数、 d は刻み幅で一定の条件を満たしつつ $d \rightarrow 0$ とする。ただしこの場合、証明がより込み入ってくるので、例えば Ait-Sahalia(1996)¹¹ を参照されたい。

¹¹Ait-Sahalia, Y. (1996), " The Delta method for Nonparametric Kernel Functionals," Department of Economics, Princeton University.

付論 :非線形計量分析の為の統計的漸近理論

2005年4月

概要

この付論は Amemiya 教授の講義で用いられる統計的漸近理論の基礎的内容を Amemiya(1985)¹ の3章・4章の一部分に基づいてまとめたものである。連続講義に先立ち、国友直人が解説したが、定理の証明については Amemiya (1985)、あるいはしかるべき文献を参照されたい。

1 大標本理論

本節では非線形計量経済分析の基礎となる確率論の内容を簡潔にまとめた。内容の多くは学部上級・大学院初級レベルの確率論の標準的内容である。

1.1 確率変数とその確率分布

1.1.1 確率変数

定義 1.1.1 (σ -加法族) 標本空間 Ω の部分集合族 \mathcal{F} が、次の3つの条件を満たすとき、 σ -加法族と云う。

- (i) $\Omega \in \mathcal{F}$
- (ii) $E \in \mathcal{F} \Rightarrow \bar{E} \in \mathcal{F}$ (\bar{E} は E の Ω に関する補集合)
- (iii) $E_j \in \mathcal{F}, j = 1, 2, \dots \Rightarrow \cup_j^\infty E_j \in \mathcal{F}$

定義 1.1.2 (確率測度) σ -加法族 \mathcal{F} 上で定義された実数値関数 $P(\cdot)$ が次の条件を満たすとき、 A 上の確率測度と云う。

- (i) $E \in \mathcal{F} \Rightarrow P(E) \geq 0$
- (ii) $P(\Omega) = 1$
- (iii) $E_i \cap E_j = \emptyset (i \neq j) \Rightarrow P(\cup_j E_j) = \sum_j P(E_j)$

定義 1.1.3 (確率空間) 標本空間 Ω 、 Ω 上の σ -加法族 \mathcal{A} 、及び \mathcal{A} 上の確率測度 $P(\cdot)$ が与えられた時、組 (Ω, \mathcal{F}, P) を確率空間と云う。

定義 1.1.4 (確率変数) 確率空間 (Ω, \mathcal{F}, P) において、 Ω 上の実数値関数 $X(\cdot)$ が任意の実数 x について、

$$\{\omega | X(\omega) < x\} \in \mathcal{F}$$

を満たす時、 X を確率変数と云う。

¹Amemiya, T. (1985), *Advanced Econometrics*, Blackwell.

1.1.2 確率分布

定義 1.1.5 (確率分布) 確率変数 $X(\omega)$ の確率分布関数 $F(x)$ を

$$F(x) = P(\omega | X(\omega) < x)$$

を満たす実数値関数で定義する。

この定義より確率分布関数 $F(\cdot)$ は以下の性質を持つことがわかる。

- (i) $F(-\infty) = 0$,
- (ii) $F(+\infty) = 1$,
- (iii) 単調非減少かつ左連続 ($\lim_{y \nearrow x} F(y) = F(x)$) 関数。

なお、同様に \mathbf{R}^d ($d \geq 1$) 上の確率分布が定義される。

Riemann-Stieltjes 積分

X を分布 F を持つ確率変数、ボレル可測な関数 $h(\cdot)$ を用いて $Y = h(X)$ と定義する。この時、 Y の期待値は次のように定義する。区間 $[a, b]$ を $a = x_0 < x_1 < \dots < x_n = b$ と n 区間に区切り、 x_i^* を区間 $[x_i, x_{i+1}]$ の任意の点とする。そこで

$$S_n = \sum_{i=0}^{n-1} h(x_i^*) [F(x_{i+1}) - F(x_i)],$$

の極限 ($n \rightarrow \infty$) により積分が定義できる。またより一般には

$$E[Y] = \int_{-\infty}^{\infty} h(x) dF(x) = \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \int_a^b h(x) dF(x)$$

で定める。このとき連続型確率分布と離散型確率分布についてはそれぞれ

$$E[h(X)] = \begin{cases} \int_{-\infty}^{\infty} h(x) f(x) dx & \text{if } \frac{dF(x)}{dx} = f(x) \\ \sum_i h(c_i) p_i & \end{cases}$$

となる。なお、非線形計量経済モデルでは離散分布と連続分布の混合分布も頻繁に登場することに注意する必要がある。

1.2 いろいろな収束

定義 1.2.1 (確率収束) 任意の $\epsilon > 0$ に対して、

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

が成り立つとき、 $\{X_n\}$ は確率変数 X へ確率収束すると云い、 $X_n \xrightarrow{P} X$ と表す。

定義 1.2.2 (平均二乗収束)

$$\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0$$

が成り立つ時、 $\{X_n\}$ は確率変数 X へ平均二乗収束すると云い、 $X_n \xrightarrow{m} X$ と表す。

定義 1.2.3 (分布収束) 確率変数列 X_n の確率分布 F_n が X の確率分布 F の全ての連続点で F に収束する時、 $\{X_n\}$ は確率変数 X へ分布収束すると云い、 $X_n \xrightarrow{d} X$ と表す。

ここで次のような $F_n(\cdot)$ を考えてみよう。

$$F_n(x) = \begin{cases} 0, & x < \alpha - \frac{1}{n} \\ \frac{n}{2}(x - \alpha + \frac{1}{n}) & \alpha - \frac{1}{n} \leq x \leq \alpha + \frac{1}{n} \\ 1 & \alpha + \frac{1}{n} < x \end{cases}$$

この時極限 $\lim_{n \rightarrow \infty} F_n$ は α で左連続ではなく、従って分布関数ではない。しかし、分布収束の定義における「全ての連続点で」という条件のゆえに、分布 F_n をもつ確率変数は、 α に確率 1 を持つ確率変数に分布収束すると見なすことができる。

定義 1.2.4 (概収束)

$$P(\omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1$$

が成り立つ時、 $\{X_n\}$ は確率変数 X へ概収束すると云い、 $X_n \xrightarrow{a.s.} X$ と表す。

定理 1.2.1 (Chebyshev) $EX_n^2 \rightarrow 0 \Rightarrow X_n \xrightarrow{p} 0$

この定理より、すぐに $X_n \xrightarrow{m} 0 \Rightarrow X_n \xrightarrow{p} 0$ となることが分かる。

定理 1.2.2 $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

定理 1.2.3 $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X$

一般には定理 1.2.2 の逆は成り立たないが、 X が定数となる場合に限っては成り立つ。

定理 1.2.4 $X_n \xrightarrow{d} \alpha \Rightarrow X_n \xrightarrow{p} \alpha$ (α : 定数)

定理 1.2.3 の逆は成り立たない。反例として以下のような例を挙げることができる。 $\Omega = [0, 1]$ 、 $A = [0, 1]$ 上のルベーグ可測な集合族、 $P =$ ルベーグ測度 (すなわち、区間の長さ) とし、次のように確率変数 $X_n(\omega)$ を定義する。

$$\begin{aligned} X_1(\omega) &= 1 \text{ for } 0 \leq \omega \leq 1 \\ X_2(\omega) &= 1 \text{ for } 0 \leq \omega \leq \frac{1}{2} \\ &= 0 \text{ elsewhere} \\ X_3(\omega) &= 1 \text{ for } \frac{1}{2} \leq \omega \leq \frac{1}{2} + \frac{1}{3} \\ &= 0 \text{ elsewhere} \\ X_4(\omega) &= 1 \text{ for } 0 \leq \omega \leq \frac{1}{12} \text{ and } \frac{1}{2} + \frac{1}{3} \leq \omega \leq 1 \\ &= 0 \text{ elsewhere} \\ X_5(\omega) &= 1 \text{ for } \frac{1}{12} \leq \omega \leq \frac{1}{12} + \frac{1}{5} \\ &= 0 \text{ elsewhere} \\ &\vdots \end{aligned}$$

このとき、任意の $1 > \epsilon > 0$ に対して、

$$P(|X_n| > \epsilon) = \frac{1}{n}$$

より、 $X_n \xrightarrow{p} 0$ であるが、 $P(\omega | \lim X_n(\omega) = 0) = 0$ となることが分かる。

次のよく知られた3つの定理は、標本数が大きくなる時に推定量や検定統計量の漸近的性質を導く際に特に有用である。

定理 1.2.5 (Mann and Wald) \mathbf{X}_n と \mathbf{X} を k 次元の確率変数ベクトルとし、 $g(\cdot)$ をその不連続点 E が閉集合かつ $P(\mathbf{X} \in E) = 0$ となるような \mathbf{R}^k から \mathbf{R} への関数とする。この時、 $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ ならば $g(\mathbf{X}_n) \xrightarrow{d} g(\mathbf{X})$ が成り立つ。

定理 1.2.6 \mathbf{X}_n を固定された有限次元の確率変数ベクトルとし、 $g(\cdot)$ を定数ベクトルの点 α で連続な実数値関数とする。この時、 $\mathbf{X}_n \xrightarrow{p} \alpha$ ならば $g(\mathbf{X}_n) \xrightarrow{p} g(\alpha)$ が成り立つ。

定理 1.2.7 (Slutsky) $X_n \xrightarrow{d} X$ かつ $Y_n \xrightarrow{p} \alpha$ ならば、

(i) $X_n + Y_n \xrightarrow{d} X + \alpha$,

(ii) $X_n Y_n \xrightarrow{d} \alpha X$,

(iii) $(X_n/Y_n) \xrightarrow{d} X/\alpha$ (ただし $\alpha \neq 0$)。

定義 1.2.5 $\{X_n\}$ を確率変数の列、 $\{a_n\}$ を正実数の列とする。この時、 $\text{plim}_{n \rightarrow \infty} a_n^{-1} X_n = 0$ ならば $X_n = o(a_n)$ 、また全ての n について $\forall \epsilon > 0, \exists M_\epsilon, P(a_n^{-1} |X_n| \leq M_\epsilon) \geq 1 - \epsilon$ ならば $X_n = O(a_n)$ と表現する。

1.3 大数の法則と中心極限定理

確率変数列 $\{X_t\}$ に対して、標本平均を $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ により定義する。

$$\bar{X}_n - E\bar{X}_n \xrightarrow{a.s.} 0$$

が成り立つ時、 $\{X_t\}$ は、大数の強法則 (strong law of large numbers) に従うと云う。また、

$$\bar{X}_n - E\bar{X}_n \xrightarrow{p} 0$$

が成り立つ時、 $\{X_t\}$ は、大数の弱法則 (weak law of large numbers) に従うと云う。ここでは、2つの基本的な強法則を挙げておく。

定理 1.3.1 (Kolmogorov LLN 1) $\{X_t\}$ は独立で、有限の分散 $V(X_t) = \sigma_t^2$ を持つとする。この時、 $\sum_1^\infty \sigma_t^2/t^2 < +\infty$ ならば、 $\bar{X}_n - E\bar{X}_n \xrightarrow{a.s.} 0$ が成り立つ。

定理 1.3.2 (Kolmogorov LLN 2) $X_t \sim i.i.d.$ の時、 $\bar{X}_n - E\bar{X}_n \xrightarrow{a.s.} 0$ となるための必要十分条件は、 $E(X_t)$ が存在し、 $E(X_t) = \mu$ となることである。

$Z_n = [V(\bar{X}_n)]^{-1/2}[\bar{X}_n - E(\bar{X}_n)]$ がある条件の下で標準正規分布 $N(0, 1)$ に分布収束するという定理を一般に中心極限定理 (central limit theorem) と呼ぶ。ここでは、3つのよく知られた中心極限定理を挙げておく。

定義 1.3.1 確率変数 X の特性関数は、 $i^2 = -1$ として $E[e^{i\lambda X}]$ で定義する。

定理 1.3.3 任意の λ について $E[\exp(i\lambda X_n)] \rightarrow E[\exp(i\lambda X)]$ で、かつ $E[\exp(i\lambda X)]$ が $\lambda = 0$ で連続ならば、 $X_n \xrightarrow{d} X$ となる。

定理 1.3.4 (Lindeberg-Levy CLT) 確率変数列 $\{X_t\}$ が *i.i.d.* (互いに独立に同一分布にしたがい) かつ、 $E(X_t) = \mu, V(X_t) = \sigma^2$ を満たすならば、 $Z_n \xrightarrow{d} N(0, 1)$ となる。

定理 1.3.5 (Liapounov CLT) $\{X_t\}$ は独立確率変数列で、 $E(X_t) = \mu_t, V(X_t) = \sigma_t^2, E[|X_t - \mu_t|^3] = m_{3t}$ とする。この時、

$$\lim_{n \rightarrow \infty} \left[\sum_{t=1}^n \sigma_t^2 \right]^{-1/2} \left[\sum_{t=1}^n m_{3t} \right]^{1/3} = 0,$$

を満たすならば、 $Z_n \xrightarrow{d} N(0, 1)$ となる。

定理 1.3.6 (Lindeberg-Feller CLT) $\{X_t\}$ は独立確率変数列で、それぞれ分布 $\{F_t\}$ を持ち、 $E(X_t) = \mu_t, V(X_t) = \sigma_t^2$ とする。また $C_n = (\sum_{t=1}^n \sigma_t^2)^{1/2}$ と定義する。この時、任意の $\epsilon > 0$ についてリンドバーグ条件

$$\lim_{n \rightarrow \infty} \frac{1}{C_n^2} \sum_{t=1}^n \int_{|X - \mu_t| \geq \epsilon C_n} (x - \mu_t)^2 dF_t = 0$$

を満足するならば、 $Z_n \xrightarrow{d} N(0, 1)$ となる。

なお、確率変数列の期待値の収束と分布収束は必ずしも対応していない。このことは次の例を見れば明らかであろう。

$$X_n = \begin{cases} 0 & \frac{n-1}{n} \\ n^2 & \frac{1}{n} \end{cases}$$

とすると $EX_n \rightarrow +\infty$ となる。

2 極値推定量の漸近的性質

非線形計量モデルの統計的推測では標本から求められる尤度関数は複雑になることが多いので計量分析においてもっとも関心がある母数に関する推定量は簡単に明示できない場合が一般的である。非線形モデルを分析する場合には極値推定量 (extremum estimator) と呼ばれている推定法を利用することが一般的である。なお、よく知られている最尤推定量 (maximum likelihood estimator) はこのタイプの推定量の例である。

パラメータ空間上である関数を最大化、あるいは最小化することによって得られる推定量を、極値推定量と呼ぶ。ここでは、極値推定量の一致性、漸近正規性が成り立つための条件に関する定理を述べておく。

2.1 有用な一般的結果

2.1.1 一致性 (consistency)

標本ベクトル \mathbf{y} 、母数ベクトル θ 、母数ベクトルの真値 $\theta_0 \in \Theta$ (母数空間) とする。実際には標本ベクトルは独立な確率変数列 X_i より $\mathbf{y} = (X_1, \dots, X_n)'$ となることが多い。

定理 2.1.1 (global maximum) 次の仮定をおく。

(A) パラメータ空間 Θ は、 \mathbb{R}^k のコンパクトな部分集合である。

(B) $Q_T(\mathbf{y}, \theta)$ は、全ての \mathbf{y} において $\theta \in \Theta$ について連続で、かつ全ての $\theta \in \Theta$ において \mathbf{y} の可測関数。

(C) $T^{-1}Q_T(\theta)$ は $\theta \in \Theta$ において $Q(\theta)$ に一様確率収束し、かつ $Q(\theta)$ は θ_0 で唯一の大域的極大値 (global maximum) を持つとする。

このとき、

$$Q_T(\hat{\theta}_T) = \max_{\theta \in \Theta} Q_T(\theta)$$

を満たす値を $\hat{\theta}_T$ とすると、 $\hat{\theta}_T \xrightarrow{P} \theta_0$ が成り立つ。

定義 2.1.1 一様確率収束：母数ベクトル θ に依存する確率変数 $Q_T(\theta)$ が一様確率収束 (convergence in probability uniformly) するとは条件

$$\lim_{T \rightarrow \infty} P[\sup_{\theta \in \Theta} |T^{-1}Q_T(\theta) - Q(\theta)| < \epsilon] = 1 \text{ for any } \epsilon > 0$$

を意味する。

定理 2.1.2 (local maximum) 次の仮定をおく。

(A) パラメータ空間 Θ は、 \mathbb{R}^k の開部分集合。

(B) $Q_T(\mathbf{y}, \theta)$ は、全ての $\theta \in \Theta$ において \mathbf{y} の可測関数で、 $\frac{\partial Q_T}{\partial \theta}$ が存在して、 θ_0 の開近傍 $N_1(\theta_0)$ において連続。

(C) $T^{-1}Q_T(\theta)$ が $Q(\theta)$ に $\theta \in \Theta$ で一様に確率収束するような θ_0 の開近傍 $N_2(\theta_0)$ が存在し、かつ $Q(\theta)$ は θ_0 で局所的極大値 (local maximum) を持つとする。

このとき Θ_T を

$$\frac{\partial Q_T}{\partial \theta} = 0$$

の解の集合とすると、任意の $\epsilon > 0$ について

$$\lim_{n \rightarrow \infty} P[\inf_{\theta \in \Theta_T} (\theta - \theta_0)'(\theta - \theta_0) > \epsilon] = 0$$

が成り立つ。

2.1.2 漸近正規性 (asymptotic normality)

定理 2.1.3 定理 2.1.2 の仮定に加えて、次の仮定をおく。

(A) $\frac{\partial^2 Q_T}{\partial \theta \partial \theta'}$ が存在し、 θ_0 の開凸近傍において連続。

(B) $\text{plim } \theta_T^* = \theta_0$ を満たすような全ての θ_T^* について、 $T^{-1}(\frac{\partial^2 Q_T}{\partial \theta \partial \theta'})_{\theta_T^*}$ が有限の非特異行列 $A(\theta_0) = \lim E T^{-1}(\frac{\partial^2 Q_T}{\partial \theta \partial \theta'})_{\theta_0}$ に確率収束する。

(C) $T^{-1}(\frac{\partial Q_T}{\partial \theta})_{\theta_0} \rightarrow N(0, B(\theta_0))$ 、ただし $B(\theta_0) = \lim E[T^{-1}(\frac{\partial Q_T}{\partial \theta})_{\theta_0} \times (\frac{\partial Q_T}{\partial \theta})_{\theta_0}]$ である。

このとき、定理 2.1.2 で定義された Θ_T のうちで $\text{plim } \hat{\theta}_T = \theta_0$ を満たす要素の列を $\{\hat{\theta}_T\}$ とすると、

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1})$$

が成り立つ。

次の定理は、一致性と漸近正規性の条件を同時に与えるものとして、応用上は便利である。

定理 2.1.6 定理 2.1.3 の仮定 A-C に加えて、次の仮定をおく。

(A) $T^{-1}Q_T(\theta)$ は $Q(\theta)$ に θ_0 の開近傍で一様確率収束する。

(B) 定理 4.1.3 で定義された $A(\theta_0)$ が負値定符号行列。

(C) $\text{plim } T^{-1}\frac{\partial^2 Q_T}{\partial \theta \partial \theta'}$ が存在し、 θ_0 の近傍で連続。

この時、定理 2.1.2、及び定理 2.1.3 の結果が同時に成り立つ。

ここで重要な応用例として標本が i.i.d. の場合、対数尤度関数

$$\log L(\mathbf{y}, \theta) = \sum_{t=1}^T \log f(X_t, \theta)$$

を最大化する母数ベクトルの最尤推定量 (mle) を $\hat{\theta}_T$ とすると、その漸近的性質について述べておく。正則条件として

$$A(\theta_0) = -B(\theta_0) = \lim_n E\left[\frac{1}{T} \frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \middle| \theta_0\right]$$

となる場合ことが仮定する。このとき上の定理より

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N[0, -A(\theta_0)^{-1}]$$

となることが導かれる。