

最尤推定法

確率分布は既知であるが、1つの母数（パラメータ θ ）が未知なる母集団を考える。このとき確率分布は θ に依存する関数とも考えられ、これを $f(x; \theta)$ で表すことにする。この母集団から独立に n 個の標本を抽出したときのデータが x_1, x_2, \dots, x_n の値をとる確率は

$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$ と考えられる。 $L(\theta)$ を尤度関数という。

x が連続の場合は、データが $(x_1, x_1 + dx_1), (x_2, x_2 + dx_2) \cdots, (x_n, x_n + dx_n)$ の値をとる確率となり、 $L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta) dx_1 dx_2 \cdots dx_n$ であるが、主題は確率値の同次積であり、 $dx_1 dx_2 \cdots dx_n$ をすべて1単位量の積と考えて先の離散データの場合の形で代表させよう。得られたデータ x_1, x_2, \dots, x_n は既知であるから $L(\theta)$ は θ のみの関数となる。

ここで2つのパラメータ ${}_1\theta, {}_2\theta$ について、 $L({}_1\theta) > L({}_2\theta)$ であればの未知パラメータの推定値としてはデータが実現しやすい ${}_1\theta$ を選ぶのが当然である。推定値としていくつもの候補値が考えられるとき、ある候補値 θ' に対して、つねに $L(\hat{\theta}) > L(\theta')$ が成立する $\hat{\theta}$ が存在すれば $\hat{\theta}$ を最尤推定値という。

一般に母集団から得られるデータは色々な組合せがあるので、これを変数（標本変量）として書き直して、尤度関数 $L(\theta) = f(X_1; \theta) \cdot f(X_2; \theta) \cdots f(X_n; \theta)$ を最大にするような統計量 $\hat{\Theta} = \Theta(X_1, X_2, \dots, X_n)$ が存在すれば、この関数を最尤推定量といい、これの統計値

（最尤推定値）として $\hat{\theta}$ が計算される。具体的には $\partial L / \partial \theta$ による関数を求め（これを尤度方程式という）、 $\partial L / \partial \theta = 0$ により最尤推定量を求める。なお、尤度方程式から解を直接求めようとすると計算が煩雑になることから対数尤度関数 $\ln L(\theta)$ にして求めることが多い。これは対数関数は単調増加関数であり今回の極値を求める計算結果に影響しないことによる。

このように確率分布を既知として尤度関数と最尤推定量を求め、最尤推定値を計算する方法を最尤推定法という。

なお確率分布のパラメータは、例えばポアソン分布は平均値1個のみであるが、正規分布は平均と分散の2個であるように、複数の場合もあり、これを同時に推定することが必要になる。確率変数 X が k 個のパラメータ $\theta_1, \theta_2, \dots, \theta_k$ に依存する分布にしたがうとし、その *p.d.f.* を $f(x; \theta_1, \theta_2, \dots, \theta_k)$ とするとすると、観測データの組 x_1, x_2, \dots, x_n に対する最尤推定量は $\hat{\Theta}_1 = \Theta_1(X_1, X_2, \dots, X_n), \hat{\Theta}_2 = \Theta_2(X_1, X_2, \dots, X_n), \dots, \hat{\Theta}_k = \Theta_k(X_1, X_2, \dots, X_n)$ と表すことができる。

ここで、正規分布 $N(\mu, \sigma^2)$ から n 個の標本が得られたとして、正規分布のパラメータ (平均値と分散) について最尤推定量を計算してみよう。

$$N(\mu, \sigma^2) \text{ p.d.f. : } f(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x_i - \theta_1)^2}{2\theta_2}\right)$$

$$L(\theta_1, \theta_2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x_i - \theta_1)^2}{2\theta_2}\right)$$

$$\ln L(\theta_1, \theta_2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \theta_2 - \frac{\sum_i (x_i - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sum_i (x_i - \theta_1)}{\theta_2}, \quad \frac{\partial \ln L}{\partial \theta_2} = -\frac{n}{2\theta_2} + \frac{\sum_i (x_i - \theta_1)^2}{2\theta_2^2}$$

$$\frac{\partial \ln L}{\partial \hat{\theta}_1} = 0, \quad \frac{\partial \ln L}{\partial \hat{\theta}_2} = 0 \Rightarrow \frac{\sum_i (x_i - \hat{\theta}_1)}{\hat{\theta}_2} = 0, \quad -\frac{n}{2\hat{\theta}_2} + \frac{\sum_i (x_i - \hat{\theta}_1)^2}{2\hat{\theta}_2^2} = 0$$

$$\therefore \hat{\theta}_1 = \frac{\sum_i x_i}{n} = \bar{x}, \quad \hat{\theta}_2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

また、ベルヌーイ試行を n 回繰返し、注目している方の事象が x 回生起する確率分布 (二項分布又はポアソン分布) の最尤推定量を求め、両分布が等しいことを確かめよう。

1) Binominal distribution

$$P(x) = {}_n C_x p^x (1-p)^{n-x}$$

$$L(p) = {}_n C_x p^x (1-p)^{n-x}$$

$$\ln L(p) = \ln {}_n C_x + x \ln p + (n-x) \ln(1-p)$$

$$\frac{d}{dp} \{ \ln {}_n C_x + x \ln p + (n-x) \ln(1-p) \} = \frac{x}{p} - \frac{n-x}{1-p}$$

$$\frac{d \ln L(p)}{d \hat{p}} = 0, \quad \therefore \hat{p} = \frac{x}{n}$$

2) Poisson distribution

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (np = \lambda)$$

$$L(\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$\ln L(\lambda) = x \ln \lambda - \ln x! - \lambda$$

$$\frac{d}{d\lambda} \{ x \ln \lambda - \ln x! - \lambda \} = \frac{x}{\lambda} - 1$$

$$\frac{d \ln L(\lambda)}{d \hat{\lambda}} = 0, \quad \hat{\lambda} = x \Rightarrow \hat{\lambda} = n \hat{p} \quad \hat{p} = \frac{x}{n}$$

次に、単回帰モデル $\hat{Y}_i = \beta_0 + \beta_1 X_i$ ($i = 1, \dots, n$) のパラメータ β_0, β_1 の最尤推定量はどうなるであろうか。いま y_1, \dots, y_n が同一分散 σ^2 で平均 $E(Y_i) = \beta_0 + \beta_1 x_i$ をもつ正規分布にしたがう独立な確率変数 Y_1, \dots, Y_n の標本観測値であるとする。すると x_i が観測されたときに y_i が観測される確率 $f(y_i : x_i)$ は

$$f(y_i : x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right]$$

である。よって尤度関数

$$L(\beta_0, \beta_1) = \prod_{i=1}^n f(y_i : x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right] = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp\left[-\frac{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right]$$

の対数尤度関数

$$\ln L(\beta_0, \beta_1) = -\ln(2\pi)^{n/2} \sigma^n - \frac{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$$

を β_0, β_1 で偏微分した尤度方程式を 0 とおいて整理すると

$$\begin{aligned} \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i &= 0 \end{aligned}$$

となる。これは、最小二乗法により単回帰モデルを推定するとき得られた正規方程式と同じである。これは線形回帰モデルの章で誤差項が正規分布に従うことを仮定 4) で置いたが、実は最小二乗法と最尤推定法とが一致することを前提にしておいた方がいろいろと都合がよいからである。当然、重回帰モデルでも同様の結果となる。

次に、ロジットモデルについて紹介しよう。例えば、ある個人 n が通学時に家から駅まで徒歩で行くか自転車で行くかなど選択可能な交通手段が複数ある時に、選択される交通手段 i は個人 n にとって最も効用が高いものとして選択されていると考えられる。この効用 u_{in} を次のように定式化する。

$$u_{in} = v_{in} + \varepsilon_{in}$$

v_{in} は確定的な変量であり ε_{in} は確率的な変量である。 ε_{in} はガンベル分布と呼ばれる次の *c.d.f.* 及び *p.d.f.* を仮定する。これは、正規分布の曲線に似ておりモデル式で扱いやすいことによる。

$$c.d.f.: F(\varepsilon) = e^{-e^{-\omega(\varepsilon-\eta)}}, \omega > 0$$

$$p.d.f.: F'(\varepsilon) = f(\varepsilon) = \omega e^{-\omega(\varepsilon-\eta)} \cdot e^{-e^{-\omega(\varepsilon-\eta)}}$$

η : mode

ω : scale parameter

このとき ε がパラメータ (η, ω) のガンベル分布にしたがう確率変数の場合 $\alpha\varepsilon + v$ は

$$Prob(\alpha\varepsilon + v \leq \delta) = Prob(\varepsilon \leq (\delta - v) / \alpha)$$

$$F\left(\frac{\delta}{\alpha} - \frac{v}{\alpha}\right) = e^{-e^{-\omega\left(\frac{\delta}{\alpha} - \frac{v}{\alpha} - \eta\right)}} = e^{-e^{-\frac{\omega}{\alpha}(\delta - (\alpha\eta + v))}}$$

より $\delta = \alpha\varepsilon + v$ はパラメータ $(\alpha\eta + v, \omega/\alpha), \alpha > 0$ のガンベル分布となる。また $\varepsilon_1, \varepsilon_2$ がパラメータ $(\eta_1, \omega), (\eta_2, \omega)$ をもつ独立のガンベル分布にしたがうとき、同時分布関数は個々の分布関数の積つまり $F(\varepsilon_1, \varepsilon_2) = F(\varepsilon_1) \cdot F(\varepsilon_2)$ であるから

$$\frac{\partial F(\varepsilon_1, \varepsilon_2)}{\partial \varepsilon_2} = \frac{\partial F(\varepsilon_1) \cdot F(\varepsilon_2)}{\partial \varepsilon_2} = F(\varepsilon_1) \frac{\partial F(\varepsilon_2)}{\partial \varepsilon_2} = F(\varepsilon_1) f(\varepsilon_2)$$

となるので

$$F(\varepsilon^*) = Prob(\varepsilon_1 - \varepsilon_2 \leq \varepsilon^*) = Prob(\varepsilon_1 \leq \varepsilon^* + \varepsilon_2) =$$

$$\int_{\varepsilon_2=-\infty}^{\infty} \int_{\varepsilon_1=-\infty}^{\varepsilon^* + \varepsilon_2} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 = \int_{\varepsilon_2=-\infty}^{\infty} F(\varepsilon^* + \varepsilon_2) \cdot f(\varepsilon_2) d\varepsilon_2$$

$$= \int_{-\infty}^{\infty} e^{-e^{-\omega(\varepsilon^* + \varepsilon_2 - \eta_1)}} \omega e^{-\omega(\varepsilon_2 - \eta_2)} e^{-e^{-\omega(\varepsilon_2 - \eta_2)}} d\varepsilon_2 = \int_{-\infty}^{\infty} \omega e^{-\omega(\varepsilon_2 - \eta_2)} e^{-e^{-\omega\varepsilon_2} (e^{-\omega\varepsilon^* + \omega\eta_1} + e^{\omega\eta_2})} d\varepsilon_2 = *$$

$$\delta = e^{-\omega\varepsilon^* + \omega\eta_1} + e^{\omega\eta_2}$$

$$* = \int_{-\infty}^{\infty} \omega e^{-\omega(\varepsilon_2 - \eta_2)} e^{-\delta e^{-\omega\varepsilon_2}} d\varepsilon_2 = \int_{-\infty}^{\infty} \frac{1}{\delta} e^{\omega\eta_2} \left\{ \omega \delta e^{-\omega\varepsilon_2} e^{-\delta e^{-\omega\varepsilon_2}} \right\} d\varepsilon_2$$

$$= \frac{1}{\delta} e^{\omega\eta_2} \int_{-\infty}^{\infty} \left(e^{-\delta e^{-\omega\varepsilon_2}} \right)' d\varepsilon_2 = \frac{1}{\delta} e^{\omega\eta_2} \left[e^{-\delta e^{-\omega\varepsilon_2}} \right]_{-\infty}^{\infty} = \frac{1}{\delta} e^{\omega\eta_2} \left(e^0 - \frac{1}{e^{\infty}} \right) = \frac{1}{\delta} e^{\omega\eta_2}$$

$$= \frac{e^{\omega\eta_2}}{e^{-\omega\varepsilon^* + \omega\eta_1} + e^{\omega\eta_2}} = \frac{1}{1 + e^{\omega(\eta_1 - \eta_2 - \varepsilon^*)}}$$

となりロジスティック分布となる。なお $F(\varepsilon^* + \varepsilon_2)$ は ε_1 の分布関数である。また、式中の

$e^{-\delta e^{-\omega\varepsilon_2}}$ は

$$e^{-\delta e^{-\omega\varepsilon_2}} = e^{-e^{\ln \delta} e^{-\omega\varepsilon_2}} = e^{-e^{-\omega\left(\varepsilon_2 - \frac{\ln \delta}{\omega}\right)}}$$

と変形されるので、パラメータ $(\ln \delta / \omega, \omega)$ の分布関数となることが分かる。

以上のガンベル分布の性質によるロジットモデルを導こう。個人 n が対象 i を選択する確率

P_{in} は、 i 以外のどの対象を選択する効用 u_{jn} よりも効用 u_{in} が大きいときである。つまり

$$\begin{aligned} P_{in} &= \text{Prob}(u_{in} > u_{jn}; i \neq j) = \text{Prob}(v_{in} + \varepsilon_{in} > v_{jn} + \varepsilon_{jn}; i \neq j) \\ &= \text{Prob}(\varepsilon_{jn} < v_{in} - v_{jn} + \varepsilon_{in}; i \neq j) \end{aligned}$$

ここで個人 n が 1 番目の対象 ($i \equiv 1$) を選択したとすると次のようになる。

$$\begin{aligned} P_{1n} &= \int_{\varepsilon_{1n}=-\infty}^{\infty} \int_{\varepsilon_{2n}=-\infty}^{v_{1n}-v_{2n}+\varepsilon_{1n}} \cdots \int_{\varepsilon_{Jn}=-\infty}^{v_{1n}-v_{Jn}+\varepsilon_{1n}} f(\varepsilon_{1n}, \varepsilon_{2n}, \dots, \varepsilon_{Jn}) d\varepsilon_{1n} d\varepsilon_{2n} \cdots d\varepsilon_{Jn} \\ &= \int_{\varepsilon_{1n}=-\infty}^{\infty} F_1(\varepsilon_{1n}, v_{1n} - v_{2n} + \varepsilon_{1n}, \dots, v_{1n} - v_{Jn} + \varepsilon_{1n}) d\varepsilon_{1n} = * \end{aligned}$$

先でもみたように、同時分布関数 $F(\varepsilon_{1n}, \dots, \varepsilon_{Jn})$ は ε_{in} が相互に独立なので

$$\begin{aligned} F(\varepsilon_{1n}, \dots, \varepsilon_{Jn}) &= F(\varepsilon_{1n}) \cdots F(\varepsilon_{Jn}) = \prod_{j=1}^{Jn} F(\varepsilon_{jn}) = \prod_{j=1}^{Jn} e^{-e^{-\varepsilon_{jn}}} = \exp\left(-\sum_{j=1}^{Jn} e^{-\varepsilon_{jn}}\right) \\ \frac{\partial F(\varepsilon_{1n}, \dots, \varepsilon_{Jn})}{\partial \varepsilon_{1n}} &= F_1(\varepsilon_{1n}, \dots, \varepsilon_{Jn}) = \frac{\partial F(\varepsilon_{1n})}{\partial \varepsilon_{1n}} F(\varepsilon_{2n}) \cdots F(\varepsilon_{Jn}) = e^{-\varepsilon_{1n}} \cdot e^{-e^{-\varepsilon_{1n}}} \prod_{j=2}^{Jn} F(\varepsilon_{jn}) = e^{-\varepsilon_{1n}} \exp\left(-\sum_{j=1}^{Jn} e^{-\varepsilon_{jn}}\right) \\ * &= \int_{-\infty}^{\infty} e^{-\varepsilon_{1n}} \exp\left(-\sum_{j=1}^{Jn} e^{-v_{1n}+v_{jn}-\varepsilon_{in}}\right) d\varepsilon_{in} = \int_{-\infty}^{\infty} e^{-\varepsilon_{1n}} \exp\left(-e^{-\varepsilon_{1n}} \left(\sum_{j=1}^{Jn} e^{v_{jn}-v_{1n}}\right)\right) d\varepsilon_{in} = ** \end{aligned}$$

ここで $\alpha = \sum_{j=1}^{Jn} e^{v_{jn}-v_{1n}}$ とすると

$$** = \frac{1}{\alpha} \int_{-\infty}^{\infty} \exp(-ae^{-\varepsilon_{in}}) \cdot ae^{-\varepsilon_{in}} d\varepsilon_{in} = \frac{1}{\alpha} [\exp(-ae^{-\varepsilon_{in}})]_{-\infty}^{\infty} = \frac{1}{\alpha} = \frac{v_{1n}}{\sum_{j=1}^{Jn} e^{v_{jn}}}$$

よって、 $P_{in} = \frac{v_{in}}{\sum_{j=1}^{Jn} e^{v_{jn}}}$, $\sum_{j=1}^{Jn} P_{jn} = 1$ である。しかるに確定値の効用関数を線形モデル式などで作

成し、モデルのパラメータが推定できれば各選択肢の選択確率が求められる。

これを 2 肢選択の場合を求めてみよう。個人 n が選択肢 i を選択する確率 P_{in} は、選択した

ものを 1、しない方を 0 とした $\delta_{in} = 1 \text{ or } 0$ をとるものとして $P_{in} = P_{1n}^{\delta_{1n}} P_{2n}^{\delta_{2n}}$ と書ける。よっ

て尤度関数は $L = \prod_{n=1}^N P_{1n}^{\delta_{1n}} P_{2n}^{\delta_{2n}}$ となり、対数尤度関数は $\ln L = \sum_{n=1}^N (\delta_{1n} \ln P_{1n} + \delta_{2n} \ln P_{2n})$ とな

る。効用関数を $v_{in} = \sum_{k=1}^K \theta_k X_{ink}$ と仮定すると、

$$\ln L = \sum_{n=1}^N \left\{ \delta_{1n} \ln \left(\frac{1}{1 + e^{-\sum_{k=1}^K \theta_k (X_{1nk} - X_{2nk})}} \right) + \delta_{2n} \ln n \left(\frac{-\sum_{k=1}^K \theta_k (X_{1nk} - X_{2nk})}{1 + e^{-\sum_{k=1}^K \theta_k (X_{1nk} - X_{2nk})}} \right) \right\} \text{となる。}$$

これを尤度方程式

$$\nabla \ln L = \left[\frac{\partial \ln L}{\partial \theta_1} \quad \dots \quad \frac{\partial \ln L}{\partial \theta_K} \right]^t$$

を 0 とおいて解くことになるが、一意に解くことが困難であるため近似解法をにより解くことになる。