最尤法前半

このレジュメは全然閉じていないので、教科書参照前提です。

1 generalized linear model

$$V(\frac{y_i}{n}) = \frac{p_i(1-p_i)}{n} \tag{1}$$

(1) は分散が x_i に依存することを示す。これは最小二乗法の前提である分散の一様性に反する*1。

一般線形モデルとは期待値の関数に対するモデルを仮定したものである。これまでの線形モデルでは、期待値そのものに対するモデルを仮定していた。以下に具体例で違いを確認する。

(これまでの) 線形モデル：$\hat{\mu},\ \hat{\alpha}$

一般線形モデル：$ln\frac{p}{1-p},\ \Phi^{-1}(p),\ lnp_{ij}$

2 maximum likelyhood method

正規線形モデルの例

$$y_i = \mu + \epsilon_i,\ i = 1, \cdots, n \tag{2}$$

最尤推定量は以下の偏微分計算によって求められる。まず期待値は

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\mu} = -\frac{1}{2\sigma^2}\Sigma 2(y_i-\mu)(-1) = \frac{1}{\sigma^2}\Sigma(y_i-\mu) = \frac{1}{\sigma^2}(n\bar{y}-n\mu) \tag{3}$$

(3) を 0 にする μ を $\tilde{\mu}$ とすると $\tilde{\mu} = \bar{y}$ なることがわかる。次に分散は

$$\frac{\partial lnL(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\Sigma(y_i-\tilde{\mu})^2 \tag{4}$$

(4) を 0 にする σ を $\tilde{\sigma}$ とすると $\tilde{\sigma}^2 = \frac{1}{n}\Sigma(y_i-\tilde{\mu})^2$ なることがわかる。一般に未知母数ベクトルを $\theta = (\theta_1, \cdots, \theta_p)'$ としたとき、最尤推定量は $\tilde{\theta}$ 以下の尤度方程式

$$\frac{\partial}{\partial\theta}lnL(\theta) = 0 \tag{5}$$

を満たす。

2 項分布モデルを考える。確率分布は

$${}_nC_y p^y(1-p)^{n-y} \tag{6}$$

で与えられる。このとき対数尤度を p で偏微分すると

$$\frac{\partial lnL(p)}{\partial p} = \frac{y}{n} - \frac{n-y}{1-p} \tag{7}$$

(7) を 0 にする p を $\tilde{p}$ とすると $\tilde{p} = \frac{y}{n}$

*1 教科書 (2.5) 参照

多項分布モデルを考える。確率分布は

$$\frac{n!}{\Pi_i\Pi_j y_{ij}!}\Pi_i\Pi_j p_{ij}^{y_{ij}}, p_{..}=1 \tag{8}$$

で与えられる。このとき対数尤度は次のようになる。

$$ln(n!/\Pi_i^a\Pi_j^b y_{ij}!)+\Sigma_i^a\Sigma_j^b y_{ij} ln p_{ij} \tag{9}$$

ここで $\Sigma_{i,j}p_{ij}=1$ であることから、以下の拘束条件を得る。

$$p_{ab}=1-p_{11}-p_{12}-\cdots-p_{a,b-1} \tag{10}$$

(9) に (10) を代入してから、p_{11} で偏微分すると

$$\frac{\partial lnL(\mathbf{p})}{\partial p_{11}}=\frac{y_{11}}{p_{11}}-\frac{y_{ab}}{1-p_{11}-p_{12}-\cdots-p_{a,b-1}} \tag{11}$$

各 i,j で同じ操作をすることによって

$$\frac{y_{11}}{p_{11}}=\frac{y_{12}}{p_{12}}=\cdots=\frac{y_{ab}}{p_{ab}} \tag{12}$$

が得られ、ここから明らかに以下の式が成立する。

$$\frac{\Sigma_{i,j}y_{ij}}{\Sigma_{i,j}p_{ij}}=\frac{y_{..}}{p_{..}}=\frac{n}{1} \tag{13}$$

したがって

$$\tilde{p}_{ij}=\frac{y_{ij}}{n} \tag{14}$$

が最尤推定量として得られる。

ラグランジュの未定係数法

上で扱った多項分布モデルの問題は、制約付き最小化問題である。ラグランジュの未定係数法[*2][*3]で解くこともできる。実際に多項分布モデルの問題に取り組む前に、ラグランジュの未定係数法について、2 変数関数の極値を計算する例を通じて原理や仕組みを説明することにする。

拘束条件

$$g(x,y)=0 \tag{15}$$

のもとで、2 変数関数 $f(x,y)$ の極値問題を考える。

拘束条件

(15) の両辺を全微分すると

$$dg(x,y)=(\frac{\partial g}{\partial x})dx+(\frac{\partial g}{\partial y})dy=0 \tag{16}$$

が成立する。

[*2] 物理の中では「解析力学」という分野でお勉強します。束縛条件の中で（例えば小球が斜面を転がる際に、ボールと斜面が離れない、と言った条件）質点の運動を追跡するときの常套手段です。そもそもラグランジュの未定係数法の原理や仕組みがわからないという方も多いので、このプリントで説明することにします。

[*3] とは言っても厳密な説明はしません。私は数学苦手なので勘弁してください。

停留条件

$f(x,y)$ を全微分すると

$$df(x,y) = (\frac{\partial f}{\partial x})dx + (\frac{\partial f}{\partial y})dy \tag{17}$$

となる。停留条件 (極値をとる条件) は

$$df = 0 \tag{18}$$

である。

2.0.1 拘束＋停留条件

(16) のもとで (18) を満たすには

$$\frac{\partial f}{\partial x} = \lambda\frac{\partial g}{\partial x},\ \frac{\partial f}{\partial y} = \lambda\frac{\partial g}{\partial y} \tag{19}$$

であればよく、実際 (19) を (17) に代入すると

$$df = \lambda\frac{\partial g}{\partial x}dx + \lambda\frac{\partial g}{\partial y}dy = \lambda dg = 0\ (\because (16)) \tag{20}$$

となって停留条件を満たしていることが確認できる。以上のことから拘束条件がある場合の極値問題は次の連立方程式を解くことに帰着できる。

$$\frac{\partial f}{\partial x} - \lambda\frac{\partial g}{\partial x} = 0 \tag{21}$$

$$\frac{\partial f}{\partial y} - \lambda\frac{\partial g}{\partial y} = 0 \tag{22}$$

$$g(x,y) = 0 \tag{23}$$

さらに変形すると

$$\frac{\partial}{\partial x}(f - \lambda g) = 0 \tag{24}$$

$$\frac{\partial}{\partial y}(f - \lambda g) = 0 \tag{25}$$

$$g(x,y) = 0 \tag{26}$$

つまり、注目している関数 f に、拘束条件 g の式に任意の関数 λ を掛け算したものを、足し合わせて偏微分操作をすれば極値が調べられると言っている。このように、ラグランジュの未定係数法は、拘束条件がある多変数関数の極値問題を機械的に処理できる便利な手法である[*4]。

解析力学の Euler-Lagrange 方程式

ここでは、物理においてラグランジュの未定係数法がどのように現れるのか説明する[*5]。座標関数[*6]q_1, q_2 について、次のような拘束条件がある場合の力学系を考察したい。

$$C(q_1, q_2) = 0 \tag{27}$$

*4 私が学部 2 年時の解析力学の先生は、この手法のことを「猿でもできるやり方」と呼んでいました。今考えると、そういった発言はアカハラと受け止められる可能性がありそうですね。ちなみにとてつもなく優秀な先生でした。

*5 完全に私の趣味です。これを書いている時、テレワークで接続ができず暇になってしまいました。しばし、お付き合いください。

*6 本当は一般化変数という量を扱いますが、これは統計の輪講なので、わかりやすい、というか慣れ親しんだ量を扱うことにします。

この系の作用[*7] S は、ラグランジアン[*8]を L とすると

$$S[q_1(t), q_2(t)] = \int_{t_1}^{t_2} L(q_1, q_2, \dot{q}_1, \dot{q}_2) dt \tag{28}$$

で与えられる。作用 S の変分[*9]をとると、

$$\delta S = \int_{t_1}^{t_2} (\frac{\partial L}{\partial q_1} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_1}) \delta q_1 dt + \int_{t_1}^{t_2} (\frac{\partial L}{\partial q_2} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_2}) \delta q_2 dt \tag{29}$$

となります[*10]。次に拘束条件についても変分を考えると

$$\delta C = \frac{\partial C}{\partial q_1} \delta q_1 + \frac{\partial C}{\partial q_2} \delta q_2 = 0 \tag{30}$$

が成立する。$\delta S = 0$ となるための条件は、$\lambda(t)$ を任意の関数として、以下の式を満たすことである。

$$\frac{\partial L}{\partial q_1} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_1} = -\lambda(t)\frac{\partial C}{\partial q_1},\ \frac{\partial L}{\partial q_2} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_2} = -\lambda(t)\frac{\partial C}{\partial q_2} \tag{31}$$

これを (29) に代入すると

$$\delta S = \int_{t_1}^{t_2} (-\lambda(t)\frac{\partial C}{\partial q_1}) \delta q_1 dt + \int_{t_1}^{t_2} (-\lambda(t)\frac{\partial C}{\partial q_2}) \delta q_2 dt = -\lambda(t) \int_{t_1}^{t_2} (\frac{\partial C}{\partial q_1}\delta q_1 + \frac{\partial C}{\partial q_2}\delta q_2) dt = 0\ (\because (30)) \tag{32}$$

となり、確かに $\delta S = 0$ を満たす。以上の議論から拘束条件がある場合に、力学系の問題を解きたい時には、以下の連立方程式を解くことに帰着できる。

$$\frac{\partial L}{\partial q_1} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_1} + \lambda(t)\frac{\partial C}{\partial q_1} = 0 \tag{33}$$

$$\frac{\partial L}{\partial q_2} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_2} + \lambda(t)\frac{\partial C}{\partial q_2} = 0 \tag{34}$$

$$C(q_1, q_2) = 0 \tag{35}$$

さらに $L' = L + \lambda C$ とすると

$$\frac{\partial L'}{\partial q_1} - \frac{d}{dt}\frac{\partial L'}{\partial \dot{q}_1} = 0 \tag{36}$$

$$\frac{\partial L'}{\partial q_2} - \frac{d}{dt}\frac{\partial L'}{\partial \dot{q}_2} = 0 \tag{37}$$

$$C(q_1, q_2) = 0 \tag{38}$$

が言える。これは元のラグランジアンに拘束条件を加えた新しいラグランジアンを使えば、系の物理的な振る舞いがわかるということを主張している[*11]。

[*7] 作用が何か、と問われるとバシッと答えられませんが、系というものは作用を最小化するような量を取るという原理があります。これを最小作用の原理と呼びます。この原理を使うと、なぜ 2 点を結ぶ線のうち、線分が最短距離になるのか導くこともできます。

[*8] 系の力学的な性質を決める量です。平たい言葉で表現すると、(運動エネルギー)-(ポテンシャル) になります。

[*9] 関数の微分に相当します。汎関数を相手にするときは、変分という概念になります。計算の見た目は微分と同じです。

[*10] 実はここで少し手間のかかる計算を行いますが、それは省略します。さらに暇だったらやります。$\delta q_i(t_1) = \delta q_i(t_2) = 0,\ i = 1, 2$ である (実現する運動の両端点は固定されている) ことと部分積分を使います。

[*11] 実はラグランジアンの任意性から自明な事実なのですが、それは物理の話なのでやめましょう。

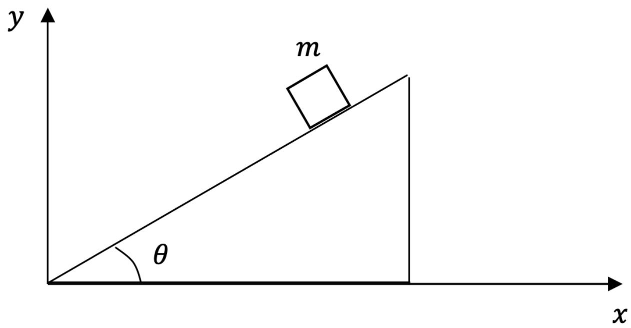


図 1　斜面を滑る小物体

簡単な力学の例題

ここでは、以下の図のように固定されていない斜面の上を小物体が滑る運動を考える[12]。ただし摩擦力は無視する。物体の位置を (x, y) と表すと、物体が斜面と常に接しているという拘束条件から

$$y = x\tan\theta \to y - x\tan\theta = 0 \tag{39}$$

という方程式が得られる。この条件のもとで新しいラグランジアン L' を書き下すと

$$L' = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgy + \lambda(y - x\tan\theta) \tag{40}$$

となる。したがって Euler-Lagrange 方程式は、まず x について

$$\frac{\partial L'}{\partial q_1} - \frac{d}{dt}\frac{\partial L'}{\partial \dot{q}_1} = 0 \to \lambda\tan\theta + \frac{d}{dt}(m\dot{x}) = 0 \to \lambda\tan\theta + m\ddot{x} = 0 \tag{41}$$

次に y について

$$\frac{\partial L'}{\partial q_2} - \frac{d}{dt}\frac{\partial L'}{\partial \dot{q}_2} = 0 \to \lambda - mg - \frac{d}{dt}(m\dot{y}) = 0 \to\to \lambda - mg - (m\ddot{y}) = 0 \tag{42}$$

ここで (39) の両辺について時間 t の二階微分をとると

$$\ddot{y} - \ddot{x}\tan\theta = 0 \tag{43}$$

ここに (41),(43) を代入すると

$$\frac{\lambda - mg}{m} - (-\frac{\lambda\tan\theta}{m})\tan\theta = 0 \ \therefore \lambda = mg\cos^2\theta \tag{44}$$

が導かれる。この λ を改めて (41),(43) に代入し直すと

$$\ddot{x} = -g\cos\theta\sin\theta,\ \ddot{y} = -g\sin^2\theta \tag{45}$$

[12] 大学受験でみんなが嫌がる問題でした。幾何学的な考察が必要になってしまう厄介さがあります。しかし、それさえもラグランジュの未定係数法を使えば、直ちに解けてしまうのです。とは言え、猿でも解けるとは思いません。

と計算できて、加速度の大きさを α とすると

$$\alpha = \sqrt{\ddot{x}^2 + \ddot{y}^2} = \sqrt{(-g\cos\theta\sin\theta)^2 + (-g\sin^2\theta)^2} = g\sin\theta \tag{46}$$

と求められ、等加速度で斜面を下ることがわかった。

脱線終わり

それでは、ラグランジュの未定係数法を独立モデルの場合に適用してみる。対数尤度は

$$lnL(\mathbf{p}) = ln(n!/\Pi_i^a \Pi_j^b y_{ij}!) + \Sigma_i^a y_{i.} lnp_{i.} + \Sigma_j^b y_{.j} lnp_{.j} \tag{47}$$

これを以下の拘束条件

$$\Sigma_i p_{i.} = 1, \Sigma_j p_{.j} = 1 \tag{48}$$

のもとで最大化することを考える。そのためには、ラグランジュ未定係数を λ_1, λ_2 として、以下の関数

$$K \equiv \Sigma_i^a y_{i.} lnp_{i.} + \Sigma_j^b y_{.j} lnp_{.j} + \lambda_1(\Sigma_i p_{i.} - 1) + \lambda_2(\Sigma_j p_{.j} - 1) \tag{49}$$

を最大化すれば良い[*13]。K を $p_{i.}, p_{.j}$ でそれぞれ偏微分して 0 とする。その後、拘束条件の式と組み合わせる。

$$\frac{\partial K}{\partial p_{i.}} = y_{i.}/p_{i.} + \lambda_1 = 0 \rightarrow p_{i.} = -y_{i.}/\lambda_1 \tag{50}$$

$$\Sigma_i p_{i.} - 1 = 0 \rightarrow \Sigma_i(-\frac{y_{i.}}{\lambda_1}) - 1 = 0 \ (\because (50)) \rightarrow -\frac{n}{\lambda_1} = 1 \rightarrow \lambda_1 = -n \tag{51}$$

$$\tilde{p}_{i.} = y_{i.}/n \tag{52}$$

同様に

$$\tilde{p}_{.j} = y_{.j}/n \tag{53}$$

今は独立モデルを考えているから

$$\tilde{p}_{ij} = y_{i.} y_{.j}/n^2 \tag{54}$$

という結果が得られる。

3 データの持つ情報量

3.1 フィッシャー情報量

原点を通る単回帰モデルは

$$\mathbf{y} = \beta\mathbf{x} + \epsilon, \ \epsilon \sim N(0, \sigma^2) \tag{55}$$

と書ける。このとき対数尤度は

$$lnL(\beta) = ln[(2\pi\sigma^2)^{-\frac{n}{2}} exp - \frac{(\mathbf{y}-\beta\mathbf{x})'(\mathbf{y}-\beta\mathbf{x})}{2\sigma^2}] = -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}-\beta\mathbf{x})'(\mathbf{y}-\beta\mathbf{x}) \tag{56}$$

これを β で偏微分する。展開すれば各項がスカラーであることは自明であり

$$\frac{\partial lnL(\beta)}{\partial\beta} = -\frac{1}{2\sigma^2}\frac{\partial}{\partial\beta}(\mathbf{y}-\beta\mathbf{x})'(\mathbf{y}-\beta\mathbf{x}) = -\frac{1}{2\sigma^2}\frac{\partial}{\partial\beta}(-2\mathbf{x}'\mathbf{y} + 2\beta\mathbf{x}'\mathbf{x}) \tag{57}$$

[*13] 言うなれば、拘束条件を付け加えた新しい関数を最大化する問題を考えたい、ということです。

が得られる。これを 0 とすると、最尤推定量

$$2\mathbf{x}'\mathbf{y} = 2\tilde{\beta}\mathbf{x}'\mathbf{x} \to \tilde{\beta} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} \tag{58}$$

が得られる。

次に対数尤度の 2 階微分を計算すると

$$\frac{\partial^2 lnL(\beta)}{\partial\beta^2} = -\frac{1}{2\sigma^2}\cdot 2\mathbf{x}\mathbf{x} = -\frac{\mathbf{x}'\mathbf{x}}{\sigma^2} \tag{59}$$

が得られ、これは β の分散を表している。すなわち推定量としての良さに対応する量となっていることがわかる。(59) の負号を外した $\frac{\mathbf{x}'\mathbf{x}}{\sigma^2}$ はフィッシャー情報量と呼ばれデータ $\mathbf{y}$ の情報量を表している。

$\mathbf{y}$ が密度関数 $f_\theta(\mathbf{y})$ に従うとき、(59) の期待値を計算すると

$$I(\theta) = -E(\frac{\partial^2 lnL(\theta)}{\partial\theta^2}) = -(\frac{\partial^2 lnf_\theta(\mathbf{y})}{\partial\theta^2}) \tag{60}$$

これもフィッシャー情報量と呼ぶ*14。行列に拡張すると

$$I_{ij}(\theta) = -E(\frac{\partial^2 lnL(\theta)}{\partial\theta_i\partial\theta_j}) = -(\frac{\partial^2 lnf_\theta(\mathbf{y})}{\partial\theta_i\partial\theta_j}) \tag{61}$$

フィッシャー情報行列なるものが作れる。

3.2 クラメール・ラオの不等式

ここでは、不偏推定量の分散の下限を与える不等式を作る。$t(\mathbf{y})$ が θ の不偏推定量であるということは

$$\int t(\mathbf{y})f_\theta(\mathbf{y})d\mathbf{y} = \theta \tag{62}$$

であると定義される。なお、積分は全空間について行う。(62) の両辺を θ で偏微分すると

$$\int t(\mathbf{y})\partial f_\theta(\mathbf{y})/\partial\theta d\mathbf{y} = 1 \tag{63}$$

が得られる。次に

$$\int f_\theta(\mathbf{y})d\mathbf{y} = 1 \tag{64}$$

であることから

$$\int \partial f_\theta(\mathbf{y})/\partial\theta d\mathbf{y} = 0 \tag{65}$$

が得られる。(63) と (65) を合わせて

$$\int (t(\mathbf{y}) - \theta)\partial f_\theta(\mathbf{y})/\partial\theta d\mathbf{y} = 1 \tag{66}$$

となる。ここで

$$\frac{\partial lnf_\theta(\mathbf{y})}{\partial\theta} = \frac{\frac{\partial f_\theta(\mathbf{y})}{\partial\theta}}{f_\theta(\mathbf{y})} \to \frac{\partial f_\theta(\mathbf{y})}{\partial\theta} = f_\theta(\mathbf{y})\frac{\partial lnf_\theta(\mathbf{y})}{\partial\theta} \tag{67}$$

*14 ここはよくわかりませんでした。上では負号付きのものをフィッシャー情報量と定義していたのですから。

であること[*15]を利用すると (66) は

$$\int (t(\mathbf{y})-\theta)(\partial lnf_\theta(\mathbf{y})/\partial\theta)f_\theta(\mathbf{y})d\mathbf{y}=1 \rightarrow E((t(\mathbf{y})-\theta)(\partial lnf_\theta(\mathbf{y})/\partial\theta))=1 \tag{68}$$

と変形できる。ところで

$$\int (t(\mathbf{y})-\theta)f_\theta(\mathbf{y})d\mathbf{y}=\int t(\mathbf{y})f_\theta(\mathbf{y})d\mathbf{y}-\int \theta f_\theta(\mathbf{y})d\mathbf{y}=\theta-\theta=0 \tag{69}$$

である。共分散の表式は

$$Cov(t(\mathbf{y})-\theta,\partial lnf_\theta(\mathbf{y})/\partial\theta)=E((t(\mathbf{y})-\theta)\partial lnf_\theta(\mathbf{y})/\partial\theta)-E(t(\mathbf{y})-\theta)E(\partial lnf_\theta(\mathbf{y})/\partial\theta) \tag{70}$$

で与えられて、ここに (69) を適用すると

$$E((t(\mathbf{y})-\theta)(\partial lnf_\theta(\mathbf{y})/\partial\theta))=Cov(t(\mathbf{y})-\theta,\partial lnf_\theta(\mathbf{y})/\partial\theta) \tag{71}$$

と書き直すことができる。相関係数が一般的に 1 以下であることを利用すると

$$\frac{(Cov(t(\mathbf{y})-\theta,\partial lnf_\theta(\mathbf{y})/\partial\theta))^2}{V(t(\mathbf{y}))\cdot V(\partial lnf_\theta(\mathbf{y})/\partial\theta)}\leq 1 \tag{72}$$

これと (68)、(71) を合わせると

$$\frac{1^2}{V(t(\mathbf{y}))\cdot V(\partial lnf_\theta(\mathbf{y})/\partial\theta)}\leq 1 \rightarrow V(t(\mathbf{y}))\geq 1/V(\partial lnf_\theta(\mathbf{y})/\partial\theta) \tag{73}$$

が得られた。次に、上式の右辺を変形していきフィッシャー情報量の表式に直せることを示す。(73) の右辺の分母について

$$V(\partial lnf_\theta(\mathbf{y})/\partial\theta)=E((\partial lnf_\theta(\mathbf{y})/\partial\theta)^2)-E(\partial lnf_\theta(\mathbf{y})/\partial\theta)^2=E((\partial lnf_\theta(\mathbf{y})/\partial\theta)^2)\ (\because (65),(67)) \tag{74}$$

であることがわかる。フィッシャーの情報量は対数尤度の 2 階微分に現れる量であったから

$$\frac{\partial^2 lnf_\theta(\mathbf{y})}{\partial\theta^2}=\frac{\partial^2 f_\theta(\mathbf{y})}{\partial\theta^2}\cdot\frac{1}{f_\theta(\mathbf{y})}-(\frac{\partial f_\theta(\mathbf{y})/\partial\theta}{f_\theta(\mathbf{y})})^2 \tag{75}$$

両辺に期待値をかますと、(65) より右辺第一項は 0 となり、第二項は対数を使った形に変形できるので

$$-E(\frac{\partial^2 lnf_\theta(\mathbf{y})}{\partial\theta^2})=E((\frac{\partial lnf_\theta(\mathbf{y})}{\partial\theta})^2) \tag{76}$$

が導かれ、これは (73) の右辺を与えている[*16]。同時にフィッシャー情報量の別の形を与えている。以上の議論から、クラメール・ラオの不等式が

$$V(t(\mathbf{y}))\geq\frac{1}{I(\theta)} \tag{77}$$

が成立することがわかった。クラメール・ラオの不等式の下限を達成する不偏推定量を有効推定量と呼ぶ。

例の計算は省略[*17]。

[*15] この変形は、もはや覚えてしまいましょう。

[*16] (76) の右辺について、教科書の書き方は非常にミスリーディングに見えました。気持ち悪かったので、私は表現を変えています。

[*17] できていると信じます。もちろん目でフォローはしますが。

4.4.1

【確率収束】

以前大数の弱法則で少し出てきた。

$X_1, X_2, \cdots, X_n$ が i.i.d. で平均と分散が存在するとき、標本平均 $\bar{X}$ は次が成り立つ。

$$\lim_{n \to \infty} P\{|\bar{X} - \mu| > \varepsilon\} = 0$$

$\bar{X}$ は μ に確率収束するという。$\bar{X} \to \mu$ in P と書く。

【概収束】

大数の強法則として知られているが、

$X_1, X_2, \cdots, X_n$ が i.i.d. で平均が存在するとき、

$$P\{\lim_{n \to \infty} \bar{X} = \mu\} = 1$$

が成り立つ。$\bar{X}$ は μ に概収束する、という。

$\bar{X} \to \mu$ a.s. (almost surely)

この極限値は定数だが…

$\tilde{\theta}_n \to \theta_0$ in P
このとき、$\tilde{\theta}_n$を
一致推定量という。

今示したいことは、

$$\lim P\{|\tilde{\theta}_n - \theta_0| > \varepsilon\} = 0$$

ここで、$\tilde{\theta}_n$ は

① $$\frac{1}{n}\sum \frac{\partial \log f_\theta(y_i)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}_n} = 0$$

を満たす。ここで、左辺の $\theta = \tilde{\theta}_n$ の代わりに $\theta = \theta_0$ としたのを見たい

$$\frac{1}{n}\sum \frac{\partial \log f_\theta(y_i)}{\partial \theta} \longrightarrow E_{\theta_0}\left(\frac{\partial \log f_\theta(y)}{\partial \theta}\right) \quad \text{in } P$$

$$= \int \frac{\partial \log f_\theta(y)}{\partial \theta} \cdot f_{\theta_0}(y)\, dy$$

確率変数　真のパラメータから来る確率

ここで $\theta = \theta_0$ とすると、

② $$\frac{1}{n}\sum \frac{\partial \log f_\theta(y_i)}{\partial \theta}\bigg|_{\theta=\theta_0} \xrightarrow{n\to\infty} \int \frac{\partial \log f_\theta(y)}{\partial \theta} \cdot f_{\theta_0}(y)\, dy \bigg|_{\theta=\theta_0}$$

$$= \int \frac{\partial f_\theta(y)}{\partial \theta} \frac{1}{f_\theta(y)}\bigg|_{\theta=\theta_0} \cdot f_{\theta_0}(y)\, dy$$

$$= \int \frac{\partial f_\theta(y)}{\partial \theta}\bigg|_{\theta=\theta_0} dy$$

$$= 0 \quad (\because 4.31)$$

① $\displaystyle \frac{1}{n}\sum \frac{\partial \log f_\theta(y)}{\partial\theta}\Big|_{\theta=\tilde{\theta}} = 0$

② $\displaystyle \frac{1}{n}\sum \frac{\partial \log f_\theta(y)}{\partial\theta}\Big|_{\theta=\theta_0} \to 0 \quad \text{in } P$

②をみたす θ の解が θ_0 のみのとき.

$$\tilde{\theta} \to \theta_0 \quad \text{in } P$$

とならなければならない. (うーん…)

【 (4.51), (4.52) で何を言いたいか 】

(4.50) の式から. $n\to\infty$ のときの $\tilde{\theta}$ の分布を知りたい.

↳ $$\underline{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i} + \sqrt{n}(\tilde{\theta}-\theta_0)\frac{1}{n}\sum_{i=1}^{n} w_i = 0$$

$$\frac{1}{\sqrt{n}}\sum \frac{(z_i - E(z_i))}{\sqrt{V(z_i)}} \sim N(0,1) \quad (\because \text{中心極限定理})$$

ここで $E(z_i)=0$, $V(z_i)=I_1(\theta_0)$ より.

$$\underline{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i} = \sqrt{I_1(\theta_0)}\,\frac{1}{\sqrt{n}}\sum \frac{z_i}{\sqrt{I_1(\theta_0)}} \sim N(0, I_1(\theta_0))$$

u と置く.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i + \sqrt{n}(\tilde{\theta}-\theta_0)\underline{\frac{1}{n}\sum_{i=1}^{n} w_i} = 0$$

$\frac{1}{n}\sum w_i \to E(w_i)$ in P (∵大数の法則)

ここで $E_{\theta_0}(w_i) = E_{\theta_0}\left\{ \left. \frac{\partial^2 \log f_\theta(y_i)}{\partial \theta^2} \right|_{\theta=\theta^*} \right\}$

$$\fallingdotseq E_{\theta_0}\left\{ \left. \frac{\partial^2 \log f_\theta(y_i)}{\partial \theta^2} \right|_{\theta=\theta_0} \right\} \quad (∵n大)$$

$$= -I_1(\theta_0)$$

$\frac{1}{n}\sum w_i \to -I_1(\theta_0)$ in P なので、十分大きなnのもと、

$\frac{1}{n}\sum w_i = -I_1(\theta_0)$ (つまり定数)

として扱う。

2つ合わせて、

$$u + \sqrt{n}(\tilde{\theta}-\theta_0)(-I_1(\theta_0)) = 0$$

$$\therefore \sqrt{n}(\tilde{\theta}-\theta_0) = \frac{u}{I_1(\theta_0)}$$

$u \sim N(0, I_1(\theta_0))$ なので、

$$\tilde{\theta} = \frac{u}{\sqrt{n}\, I_1(\theta_0)} + \theta_0 \sim N\left(\theta_0, \frac{1}{n I_1(\theta_0)}\right)$$

【(4.54)と(4.55)が同値であること】

$$(*)\quad \frac{\partial^2 \log f_\theta(y)}{\partial\theta\partial\theta'} = \underbrace{\frac{\partial}{\partial\theta}\Big\{\underbrace{\frac{\partial f_\theta(y)/\partial\theta'}{f_\theta(y)}}_{\text{ベクトル}}\Big\}}_{\text{行列}}$$

$$= \frac{1}{f_\theta(y)^2}\left\{\frac{\partial^2 f_\theta(y)}{\partial\theta\partial\theta'} f_\theta(y) - \frac{\partial f_\theta(y)}{\partial\theta'}\cdot\frac{\partial f_\theta(y)}{\partial\theta}\right\}$$

$$= \underbrace{\frac{\partial^2 f_\theta(y)}{\partial\theta\partial\theta'}\cdot\frac{1}{f_\theta(y)}}_{A} - \left\{\frac{\partial f_\theta(y)}{\partial\theta}\cdot\frac{1}{f_\theta(y)}\right\}\cdot\left\{\frac{\partial f_\theta(y)}{\partial\theta}\cdot\frac{1}{f_\theta(y)}\right\}'$$

ここでAの期待値は、

$$E(A) = \int \frac{\partial^2 f_\theta(y)}{\partial\theta\partial\theta'}\cdot\frac{1}{f_\theta(y)} f_\theta(y)\,dy$$

$$= \int \frac{\partial^2 f_\theta(y)}{\partial\theta\partial\theta'}\,dy$$

$$= \underset{\text{ゼロ行列}}{O}$$

$\int f_\theta(y)dy = 1$ より、

$\int \frac{\partial f_\theta(y)}{\partial\theta}dy = 0$ ゼロベクトル

$\int \frac{\partial^2 f_\theta(y)}{\partial\theta\partial\theta'}dy = O$ ゼロ行列

よって(*)の両辺期待値をとると成立。

【例 4.6 (特に見る必要ないが、解いておく)】

$$P(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

$$\frac{\partial \log L(\mu,\sigma^2)}{\partial \mu} = \frac{\partial}{\partial \mu}\left\{c - \frac{(y-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{y-\mu}{\sigma^2} \qquad\qquad (2\pi\sigma^2)^{-\frac{1}{2}}$$

$$\frac{\partial \log L(\mu,\sigma^2)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left\{c - \frac{1}{2}\log\sigma^2 - \frac{(y-\mu)^2}{2\sigma^2}\right\}$$

$$= -\frac{1}{2\sigma^2} + \frac{(y-\mu)^2}{2\sigma^4}$$

$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$ より、

$$I_1(\mu,\sigma^2) = E\left\{\left(\frac{\partial \log L(\mu,\sigma^2)}{\partial \mu}\right)\cdot\left(\frac{\partial \log L(\mu,\sigma^2)}{\partial \sigma^2}\right)'\right\}$$

$$= E\left\{\begin{pmatrix} \frac{(y-\mu)^2}{\sigma^4} & -\frac{y-\mu}{2\sigma^4} + \frac{(y-\mu)^3}{2\sigma^6} \\ -\frac{y-\mu}{2\sigma^4} + \frac{(y-\mu)^3}{2\sigma^6} & \frac{\{(y-\mu)^2-\sigma^2\}^2}{4\sigma^8} \end{pmatrix}\right\}$$

(1, 1) 成分 : $E\left\{ \dfrac{(y-\mu)^2}{\sigma^4} \right\} = \dfrac{\sigma^2}{\sigma^4} = \dfrac{1}{\sigma^2}$

(1, 2) (2, 1) 成分 :

$$E\left\{ \underbrace{-\frac{y-\mu}{2\sigma^4}}_{0} + \frac{(y-\mu)^3}{2\sigma^6} \right\} = \frac{1}{2\sigma^6} E\left\{ (y-\mu)^3 \right\}$$

$$= \frac{1}{2\sigma^6} \int_{-\infty}^{\infty} \underbrace{(y-\mu)^3 f(y)}_{\mu\text{中心に奇関数}} dy$$

$$= 0$$

(2, 2) 成分 :

$$E\left[\frac{\left\{ (y-\mu)^2 - \sigma^2 \right\}^2}{4\sigma^8} \right] = \frac{1}{4\sigma^8} \Big[E\left\{ (y-\mu)^4 \right\} - 2\sigma^2 E\left\{ (y-\mu)^2 \right\} + \sigma^4 \Big]$$

$$= \frac{1}{4\sigma^8} \left(\underbrace{3\sigma^4}_{\frac{E\{(y-\mu)^4\}}{\sigma^4} = 3 \text{ から}} - 2\sigma^4 + \sigma^4 \right)$$

$$= \frac{1}{2\sigma^4}$$

よって

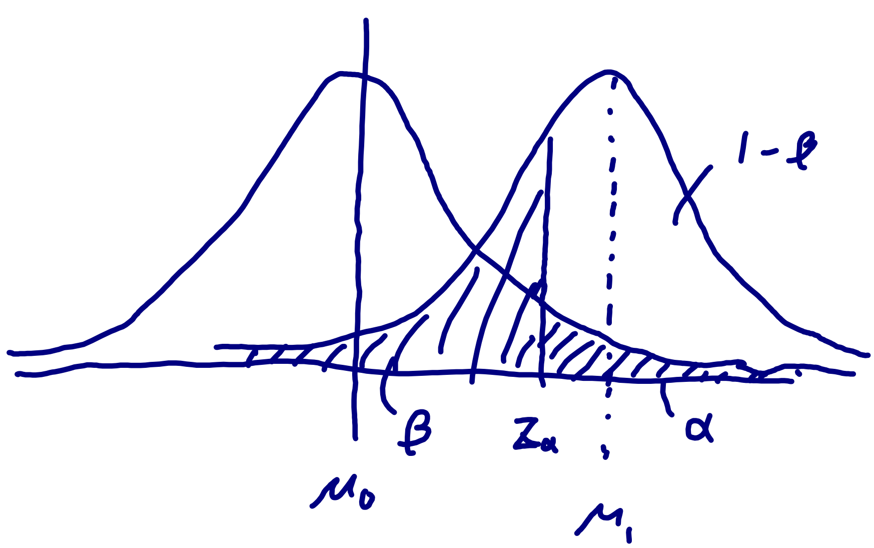
$$I_1(\mu, \sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

$$I_1(\mu,\sigma^2)^{-1} = \begin{pmatrix}\sigma^2 & 0\\ 0 & 2\sigma^4\end{pmatrix}$$

よって. $\sqrt{n}(\tilde{\theta}-\theta_0) = \sqrt{n}\begin{pmatrix}\tilde{\mu}-\mu_0\\ \tilde{\sigma}^2-\sigma_0^2\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, I_1(\mu_0,\sigma_0^2)^{-1}\right)$

$$\therefore \begin{pmatrix}\tilde{\mu}\\ \tilde{\sigma}^2\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_0\\ \sigma_0^2\end{pmatrix}, \begin{pmatrix}\frac{\sigma^2}{n} & 0\\ 0 & \frac{2\sigma^4}{n}\end{pmatrix}\right)$$

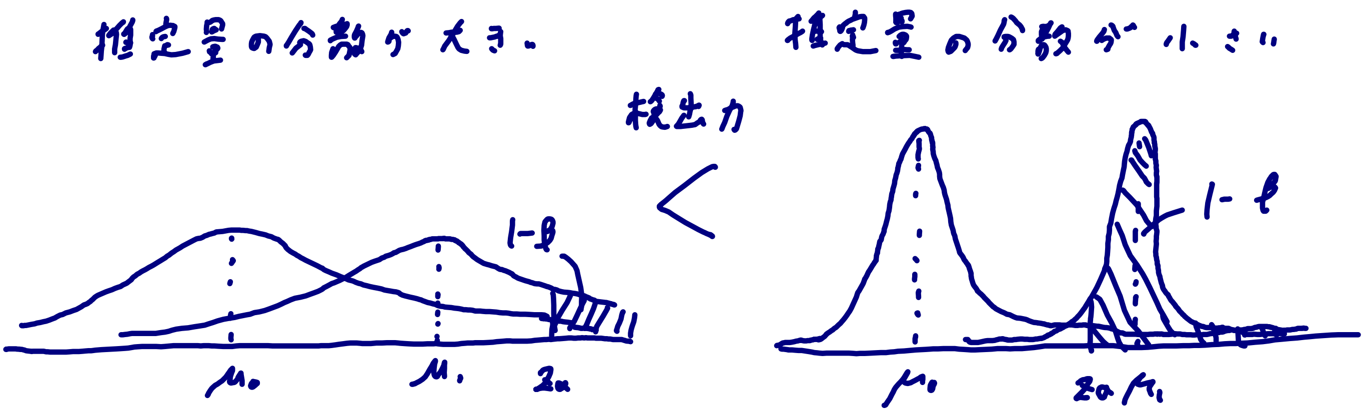
【常に最大の検出力を与える】



$\alpha = P($棄却$\mid H_0)$: 第一種の過誤 . 生産者危険

$\beta = P($採択$\mid H_1)$: 第二種の過誤 . 消費者危険

$1-\beta$: 検出力



○最尤推定量は漸近有効性から n大のときに最も分散が小さくなるので、検出力も大きくなる。

4.5.2

問 係数 $n^{-1} \cdot \partial^2 \log L(\theta)/\partial\theta^2 |_{\theta=\theta^*}$ は…… $-I_1(\theta_0)$ に確率収束
(4.47)、(4.49) (4.52) あたりと同じ。

同じ

(4.47) の第二項

$$(\tilde{\theta} - \theta_0) \sum_{i=1}^{n} \underbrace{\left. \frac{\partial^2 \log f_\theta(y_i)}{\partial \theta^2} \right|_{\theta=\theta^*}}_{w_i \text{ とおく. } (4.49)}$$

(4.49) $\dfrac{w_1 + \cdots + w_n}{n}$ は大数の法則より、$E(w_i)$ に収束する。

(4.52) $E(w_i) = -I_1(\theta_0)$

まとめると、

$$n^{-1} \partial^2 \log L(\theta)/\partial\theta^2 |_{\theta=\theta^*} = \frac{w_1 + \cdots + w_n}{n} \xrightarrow{n\text{大}} E(w_i) = -I_1(\theta_0)$$

【(4.61)】

$$2\log\frac{L(\tilde\theta)}{L(\theta_0)} = 2\log L(\tilde\theta) - 2\log L(\theta_0)$$

$$= 2\log L(\tilde\theta) - 2\log L(\tilde\theta) - \left.\frac{\partial^2 \log L(\theta)}{\partial\theta^2}\right|_{\theta=\theta^*}(\theta_0-\theta^*)^2$$

$$= -\frac{1}{n}\left.\frac{\partial^2 \log L(\theta)}{\partial\theta^2}\right|_{\theta=\theta^*}\left\{\sqrt{n}(\tilde\theta-\theta_0)\right\}^2$$

$$= I_1(\theta_0)\left\{\sqrt{n}(\tilde\theta-\theta_0)\right\}^2$$

$$= \Big\{\underbrace{\sqrt{nI_1(\theta_0)}(\tilde\theta-\theta_0)}_{\sim N(0,1)}\Big\}^2$$

P.131

$$\left(\because\ \tilde\theta \sim \left(\theta_0, \frac{1}{nI_1(\theta_0)}\right)\right)$$

P.138

【$(\bar{y}-\lambda_0)/\lambda_0$ が $n^{-\frac{1}{2}}$ のオーダーで無視できる】

$$\frac{1}{\sqrt{n}}\Big\{\underbrace{\sqrt{n}\,\frac{\bar{y}-\lambda_0}{\lambda_0}}_{\text{中心極限定理}}\Big\} = n^{-\frac{1}{2}}\cdot N(0,1)$$

【例4.11 正規分布の平均のエフィシエントスコア出してみる】

$$ES = \frac{\partial \log f_\theta(y)}{\partial \theta}$$

$$= \sum_{1}^{n} \frac{\partial}{\partial \mu} \left\{ -\frac{(y_i-\mu)^2}{2\sigma^2} + c \right\}$$

$$= n \frac{\bar{y}-\mu}{\sigma^2}$$

よって、4.64は、

$$n^{-\frac{1}{2}} \cdot n \frac{\bar{y}-\mu}{\sigma^2} > z_\alpha \sqrt{I_1(\mu_0)}$$

ここで $I_1(\mu_0) = \frac{1}{\sigma^2}$

(例4.6)

$$I_1(\mu) = -E\left(\frac{\partial^2 \log f_\mu(y)}{\partial \mu^2}\right)$$

$$= -E\left(-\frac{1}{\sigma^2}\right)$$

$$= \frac{1}{\sigma^2}$$

より、

$$\sqrt{n} \frac{\bar{y}-\mu}{\sqrt{\sigma^2}} > z_\alpha$$

これはP135例4.7と一致する

【例4.12】

$$ES = \frac{\partial \log f_\lambda(y)}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left\{ -n\lambda + y_{\cdot} \log \lambda + C \right\}$$

$$= -n + \frac{y_{\cdot}}{\lambda}$$

あとは教科書.

【(4.66)】

例4.2 の分割表の例を使うと.

$$\mathbb{P} = \begin{pmatrix} p_{11} & \cdots & p_{1b} \\ \vdots & & \\ p_{a1} & \cdots & p_{ab} \end{pmatrix} \qquad \beta = \begin{pmatrix} p_{1\cdot} \\ \vdots \\ p_{a\cdot} \\ p_{\cdot 1} \\ \vdots \\ p_{\cdot b} \end{pmatrix}$$

$$g_i(\beta) = p_{i\cdot} \begin{pmatrix} p_{\cdot 1} \\ \vdots \\ p_{\cdot b} \end{pmatrix}' \qquad \left(= (p_{i1}\ p_{i2} \cdots p_{ib}) \right)$$

として.

$$H_0 : \mathbb{P} = \begin{pmatrix} g_1(\beta) \\ \vdots \\ g_a(\beta) \end{pmatrix}$$ となる.

【4.64】

$$2\log\{L(\tilde{\theta})/L(g(\tilde{\beta}))\}$$

$$=2\{\log(L(\tilde{p}))-\log(L(\underbrace{\tilde{p}_{H_0}}_{}))\}$$

H_0下でのpの最尤推定値.

$$=2\left\{\sum_{i=1}^{a}\sum_{j=1}^{b}y_{ij}\log\tilde{p}_{ij}-\sum^{a}\sum^{b}y_{ij}\log\tilde{p}_{H_0 ij}\right\}$$

(∵ 4.10)

$$=2\left\{\sum\sum y_{ij}\log\frac{y_{ij}}{n}-\sum\sum y_{ij}\log\frac{y_{i\cdot}y_{\cdot j}}{n^2}\right\}$$

$=$ (教科書)

【(4.70)】

(4.64) $$n^{-\frac{1}{2}}\frac{\partial\log f_\theta(y)}{\partial\theta}\bigg|_{\theta=\theta_0}>z_\alpha\sqrt{I_1(\theta_0)}$$

$$n^{-\frac{1}{2}}I_1(\theta_0)^{-\frac{1}{2}}\frac{\partial\log f_\theta(y)}{\partial\theta}\bigg|_{\theta=\theta_0}>z_\alpha$$

両辺二乗して.

$$n^{-1}(I_1(\theta_0))^{-1}\left\{\frac{\partial\log f_\theta(y)}{\partial\theta}\bigg|_{\theta=\theta_0}\right\}^2>\chi^2(1)_\alpha$$

(4.70)は：mの多次元の形をしている.