7 小標本のための手法

医学研究は，大規模にデータが集められる理想的な状況でのみ行われるわけではない．ミシガン ECMO 試験のようにサンプルサイズが小さいこともある．このとき，最尤法の前提条件が満たされないため，それに代わる手法が必要になる．この章で述べられるのは小標本しか得られない状況への対処法である．統計学の文献では多くのテクニックや統計手法が提案されてきたが，ごく大まかにいえば変数変換，正確な確率計算，ペナルティ付き尤度という 3 つのアプローチに大別できる．具体的な手法として，2 値データのための信頼区間の構成方法を 3 つ紹介する．

キーワード	Agresti 法，Clopper–Pearson 法，正確な確率計算，Bayes 流の推測，ペナルティ付き尤度，変数変換，Mid-p 法
事　例	ハーバード ECMO 試験，ミシガン ECMO 試験

7.1 変数変換，正確な確率計算，ペナルティ付き尤度[*1)]

Wald 信頼区間は，汎用性が高く計算負荷が小さいため，ソフトウェアのデフォルトになっていることが多い．しかし，サンプルサイズが小さいと正規分布への近似精度が落ちる．目安として，2 値データ・計数データ・生存時間データの解析では，パラメータの数のおよそ 5 倍のイベント数が観察できない場合は，別の手法を検討すべきである．

これまで統計学の文献で提案されてきたアプローチは 3 つに大別できる．第一のアプローチは，すでに述べたパラメータ変換である．第二のアプローチと

[*1)] 紙面の都合のため，この章で紹介できた手法は 2 値データのための信頼区間の構成方法に限られている．これは小標本が問題になる典型的な状況が 2 値データだからである．それ以外のデータについては，ペナルティ付き尤度または Bayes 流の推測を応用するアプローチがよい．

して，近似は行わず，正確な確率計算によって，信頼区間を構成したり，p 値を計算したりすることもできる．ただし，パラメータの数が多い一般化線型モデルでは，このアプローチは計算負荷が大きい．そこで実践的に重要になるのはペナルティ付き尤度を用いるアプローチである．

7.2 Clopper–Pearson 信頼区間

実際のデータ解析では，2 値データの割合 π の信頼区間を求める問題によく遭遇する．このとき有用なのが正確な信頼区間（Clopper–Pearson 信頼区間）である（Clopper and Pearson 1934）．この信頼区間の下側限界は，2 項分布に従う確率変数の実現値を y，信頼係数を α とすると，2 項分布の確率関数を含む方程式

$$\sum_{i=0}^{y} \Pr(Y = i; \pi, N) = \frac{\alpha}{2}$$

を π について解くことで得られる．この方程式を直接解くのは面倒である．だが，2 項分布と F 分布の分布関数には，以下の関係があることが知られていて，これを利用することができる．

$$\sum_{i=0}^{y} \Pr(Y = i; \pi, N) = \Pr\left[F < \frac{(N - y + 1)\pi}{y(1 - \pi)}\right]$$

ここで F は，F 分布に従う確率変数を表す．これを π について解いて，Clopper–Pearson の下側限界

$$100\,(1 - \alpha)\,\%\ \text{lower limit} = \frac{y}{y + (N - y + 1)F_{\alpha/2}}$$

が導かれる．ここで，$F_{\alpha/2}$ は自由度 $2(N-y+1)$ と $2y$ を持つ F 分布の $100(1-\alpha/2)$ パーセント点である．上側限界は

$$\sum_{i=y}^{N} \Pr(Y = y; \pi, N) = \frac{\alpha}{2}$$

と定義され，同じように F 分布を参照することで計算できる．

7.3 Mid-p 法

一般に，正確な手法は，参照する確率分布が離散分布であるため，指定した

水準通りの確率計算ができず，保守的になることがある[*2)]．このような離散検定の特徴は，Mid-p 法を用いることで補正することができる（Lancaster 1949）．Mid-p 法とは，信頼区間や p 値の計算において，裾側確率 $\Pr(Y \le y)$ の代わりに，観測データが生じる 1 点の確率の半分だけ減らした，$\Pr(Y < y) + \Pr(Y = y)/2$ を用いる手法である．

Mid-p 法で補正した Clopper–Pearson 信頼区間の上側限界と下側限界は，それぞれ

$$\sum_{i=0}^{y-1} \Pr(Y = i; \pi, N) + \frac{1}{2} \Pr(Y = y; \pi, N) = \frac{\alpha}{2}$$

$$\sum_{i=y+1}^{N} \Pr(Y = i; \pi, N) + \frac{1}{2} \Pr(Y = y; \pi, N) = \frac{\alpha}{2}$$

の解として定義される．つまり，裾側確率の計算において，確率の半分つまり $\Pr(Y = y; \pi, N)/2$ だけ加えないことで，p 値を小さく（有意になりやすく）補正するのである．

7.4 事例：2 値アウトカムの臨床試験の解析 2

ミシガン ECMO 試験の次に行われた ECMO 臨床試験（ハーバード ECMO 試験）の第 2 ステージでは，ECMO により治療を受けた 20 人中 1 人が死亡した（O’Rourke, et al. 1989）．このデータから，死亡割合の 95%Wald 信頼区間を求めると

$$95\%\ \mathrm{CI} = 0.05 \pm 1.96\sqrt{0.05(1-0.05)/20} = [-0.046,\ 0.146]$$

となる．このように，2 項確率の 95%Wald 信頼区間は，[0, 1] の範囲を越えることがある．一方で，95%Clopper–Pearson 信頼区間を計算すると，正確な手法と Mid-p 法の結果はそれぞれ [0.001, 0.249] と [0.003, 0.223] となる．このように Wald 信頼区間と Clopper–Pearson 信頼区間は，サンプルサイズが小さいときかなり異なる結果を与える．

[*2)] 信頼区間が広すぎたり，検定が過度に有意になりにくかったりする傾向を保守的（conservative）という．たとえば 95%信頼区間であれば，その区間が広く，言い換えれば真値を含む確率が 95%より高いことをいう．

7.5　ペナルティ付き尤度

正確な手法よりも汎用的なアプローチとしてペナルティ付き尤度*3) がある．これらの手法は，対数尤度関数にペナルティ項（正則化項）を追加した

$$pl(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) + \text{penalty}$$

に基づいて推測を行う．Bayes 流の推測では本質的に，ペナルティ項として事前分布の対数が用いられている．それ以外にもよい性質を持つペナルティ項としてさまざまなものが考えられる．

歴史的に 2 項分布の推測のためによく用いられてきたのは，ベータ分布に由来するペナルティ項である．2 項尤度

$$l(\pi) = y\log(\pi) + (N - y)\log(1 - \pi)$$

に，ベータ分布の確率密度関数の対数を加えると

$$\begin{aligned} pl(\pi) &= y\log(\pi) + (N - y)\log(1 - \pi) + (a - 1)\log(\pi) + (b - 1)\log(1 - \pi) \\ &= (y + a - 1)\log(\pi) + (N - y + b - 1)\log(1 - \pi) \end{aligned}$$

となる．これは，2 項分布の 0 または 1 が出現した回数に，それぞれ a と b を足したときの対数尤度関数と同じものである．ベータ分布は，共役事前分布といって，ペナルティを付けた後も，関数形は 2 項尤度の形式のまま変わらないという特性がある．a と b に小さな値を加えただけで，Wald 信頼区間の性能をかなり改善することができる．Agresti（2000）によれば，$a = b = 2$ を用いた

$$\widehat{\pi} = \frac{y + 2}{N + 4}, \quad 95\%\ \text{CI} = \widehat{\pi} \pm 1.96\sqrt{\frac{\widehat{\pi}(1 - \widehat{\pi})}{N + 4}}$$

は，ほとんどあらゆる状況で 95%Wald 信頼区間よりも優れている．これを Agresti（アグレスティ）信頼区間という．また，Gart and Zweifel（1967）は，対数オッズ $\theta = \log[\pi/(1-\pi)]$ の推定においていくつかの推定量を検討した結果，$a = b = 1/2$ というわずかな補正を加えた

$$\widehat{\theta} = \log\left(\frac{y + 1/2}{N - y + 1/2}\right)$$

*3)　ペナルティ付き尤度は統計学のさまざまな分野で使われてきたテクニックである．そのため，文脈に応じて，正則化（reguralization），縮小推定（shrinkage），Bayes 推定と呼ばれることもある．

$$95\%\ \mathrm{CI} = \widehat{\theta} \pm 1.96\sqrt{\frac{1}{y+1/2} + \frac{1}{N-y+1/2}}$$

を推奨している．この推定量・標準誤差は，最尤法によるものより，サンプルサイズが増えると真値に速く収束することがわかっている．これらの手法は，a と b の選択に恣意性を感じるかもしれないが，$y = 0$ または $y = N$ のときにも計算できるという点で魅力的である．

サンプルサイズに比べてパラメータの数が多いとき，正確な手法は計算負荷が大きい．それに対して，ペナルティ付き尤度は，パラメータの数が増えたり，モデルが複雑になったりしても，Bayes 推測の枠組みで拡張することができる．特に，小規模な研究で 2 値データに一般化線型モデルを当てはめるときには，事前分布として Cauchy 分布を用いた Bayes 流のロジスティック回帰が推奨されている（Gelman，et al. 2008）．

7.6 事例：2 値アウトカムの臨床試験の解析 3

ミシガン ECMO 試験では，ECMO 群と従来療法群の死亡割合は，それぞれ 0%と 100%で，数字上は大きな差があった．しかしこの場合は，試験デザインが特殊で，サンプルサイズが非常に小さい．このようなときどのように解析すればよいだろうか．この場合は，仮説検定の性能が保証されないから，死亡割合に差がないという帰無仮説が棄却されるかどうかで結論を下すべきではない．観察された死亡割合にどれくらいの誤差があるかを表示した方が有益である．

表 7-1 に，4 通りの手法で求めた 95%信頼区間を示す．Clopper–Pearson 法，Mid-p 法，Agresti 法のいずれも，従来療法群の信頼区間の幅は非常に広い．これは，1 人しかいない従来療法群の患者が偶然によって死亡したという可能性が否定できないことを示している．

表 7-1 ミシガン ECMO 試験データにおける信頼区間の比較

	ECMO		従来療法	
	割合	95%信頼区間	割合	95%信頼区間
Wald 法	0%	推定不能	100%	推定不能
Clopper–Pearson 法	0%	0 ~ 28.5%	100%	2.5 ~ 100%
Mid-p 法	0%	0 ~ 23.8%	100%	5.0 ~ 100%
Agresti 法	0%	0 ~ 30.5%	100%	17.1 ~ 100%

演習問題

〈Agresti 信頼区間〉

問 1　対象者数 10 人の臨床試験において，2 人が死亡した．死亡割合の 95%Wald 信頼区間と 95%Agresti 信頼区間として，正しい組み合わせを選べ．ただし，数値は 3 桁で丸めてある．

(A) Wald CI = $[0, 0.448]$,　Agresti CI = $[0.05, 0.522]$
(B) Wald CI = $[-0.05, 0.448]$, Agresti CI = $[0.05, 0.522]$
(C) Wald CI = $[0.05, 0.522]$,　Agresti CI = $[0, 0.448]$
(D) Wald CI = $[0.05, 0.522]$,　Agresti CI = $[-0.05, 0.448]$