

メタアナリシスにおける正確な信頼区間

菅澤翔之助 リスク解析戦略研究センター 特任研究員

メタアナリシス

- 複数の研究の結果を統合し、より高い見地から分析する。
- ランダム化比較試験のメタアナリシスは、根拠に基づく医療(EBM: Evidence Based Medicine)において、最も質の高い根拠とされる。
- 変量効果モデルによるアプローチが一般的に用いられている。

単変量モデル

- n 個の研究があり、 y_i, σ_i^2 を i 番目の研究での効果量の推定値とその分散とする。このとき、以下の変量効果モデルを考える。

$$y_i = \theta_i + e_i, \quad \theta_i = \mu + \varepsilon_i, \quad i = 1, \dots, n.$$

ここで θ_i は i 番目の研究での真の効果量、 μ は平均効果量である。また、 e_i と ε_i は独立で $e_i \sim N(0, \sigma_i^2)$ 、 $\varepsilon_i \sim N(0, \tau^2)$ を仮定する。分散パラメータ τ^2 は未知で、研究間の不均一性を表現するパラメータである。

- 対数尤度関数

$$L(\mu, \tau^2) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n \log(\tau^2 + \sigma_i^2) - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\tau^2 + \sigma_i^2}.$$

- 尤度比検定

仮説 $H_0: \mu = \mu_0$ の尤度比検定統計量は以下のように定義される。

$$T_{\mu_0}(Y) = \max_{\mu, \tau^2} L(\mu, \tau^2) - \max_{\tau^2} L(\mu_0, \tau^2).$$

統計量 $T_{\mu_0}(Y)$ の分布は研究数 n が大きいとき χ_1^2 分布で近似できるが、 n があまり大きくないときは良い近似を与えない。(局外母数 τ^2 が原因。)

- τ^2 の最尤推定量の十分性を利用

$T_{\mu_0}(Y)|\hat{\tau}^2$ の分布は τ^2 に依存しないので、この分布で p 値を計算すれば n に依らず正確な p 値が計算できる。

Lindqvist and Taraldsen (2005)によるMonte Carlo法を利用する。

- 正確な p 値を計算するためのアルゴリズム

$\hat{\tau}^2$ を $\mu = \mu_0$ のもとでの τ^2 の最尤推定量とする。また $u_i, i = 1, \dots, n$ を互いに独立で $N(0, 1)$ に従う確率変数とし、 $U = (u_1, \dots, u_n)$ と定義する。

このとき以下の量を定義

$$\tau_*^2(U) = \left\{ \sum_{i=1}^n \frac{u_i^2}{(\hat{\tau}^2 + \sigma_i^2)^2} \right\}^{-1} \left\{ \sum_{i=1}^n \frac{\hat{\tau}^2 + \sigma_i^2(1 - u_i^2)}{(\hat{\tau}^2 + \sigma_i^2)^2} \right\},$$

$$w(U) = \left\{ \sum_{i=1}^n \frac{u_i^2}{(\hat{\tau}^2 + \sigma_i^2)^2} \right\}^{-1} \sum_{i=1}^n \frac{2\{\tau_*^2(U) + \sigma_i^2\}u_i^2 - (\hat{\tau}^2 + \sigma_i^2)}{(\hat{\tau}^2 + \sigma_i^2)^3},$$

アルゴリズム

- $\mu = \mu_0$ のもとでの τ^2 の最尤推定量 $\hat{\tau}^2$ および検定統計量 $T_{\mu_0}(Y)$ を計算。
- 十分大きな B を設定し、 $b = 1, \dots, B$ に対して、乱数 $U^{(b)}$ を生成し、 $\tau_*^2(U^{(b)})$ および $w(U^{(b)})$ を計算。さらに $Y_*^{(b)} = (y_1^{(b)}, \dots, y_n^{(b)})$, $y_i^{(b)} = \mu_0 + u_i^{(b)} \sqrt{\tau_*^2(U^{(b)}) + \sigma_i^2}$ を生成。
- 以下を計算

$$\frac{\sum_{b=1}^B I\{T_{\mu_0}(Y_*^{(b)}) \geq T_{\mu_0}(Y)\} w(U^{(b)})}{\sum_{b=1}^B w(U^{(b)})}.$$

- 正確な信頼区間

$p(\mu_0)$ を仮説 $H_0: \mu = \mu_0$ の正確な p 値とする。平均効果量 μ の正確な $100(1 - \alpha)\%$ 信頼区間は $\{\mu; p(\mu) \geq \alpha\}$ で与えられる。(上限, 下限をそれぞれ二分法などで計算。)

- 数値実験による比較

設定: $n = 3, 5, 7, 9$, $\alpha = 0.05$, $\mu = -0.80$, $\tau^2 = 0.1, 0.2$, $\sigma_i^2 \sim 0.25\chi_1^2$ ([0.009, 0.6]で切断)。

LR: 尤度比検定統計量の χ^2 近似による信頼区間

DL: DirSimonian and Laird (1986)による信頼区間

REML: 制限付き最尤法を用いた信頼区間

ECI: 提案した正確な信頼区間

2000回の繰り返しから被覆確率(CP)および平均区間長(AL)を計算。

		CP				AL			
		n	3	5	7	9	3	5	7
$\tau^2 = 0.10$	ECI	0.965	0.948	0.951	0.958	2.181	1.000	0.813	0.616
	REML	0.780	0.834	0.925	0.892	0.786	0.627	0.820	0.506
	DL	0.782	0.833	0.928	0.892	0.753	0.617	0.821	0.504
	LR	0.826	0.874	0.945	0.914	0.910	0.699	0.875	0.545
$\tau^2 = 0.20$	ECI	0.964	0.944	0.948	0.955	2.644	1.280	1.021	0.791
	REML	0.770	0.846	0.915	0.892	1.019	0.831	0.926	0.664
	DL	0.762	0.836	0.914	0.892	0.988	0.814	0.925	0.659
	LR	0.803	0.877	0.934	0.914	1.176	0.922	0.992	0.709

提案手法は n が小さいときでも正確な被覆確率を与えている。他の手法は基本的に漸近近似を用いているので n が小さいときのパフォーマンスは良くない。

- 実データへの適用

Teo et al. (1991)で扱われた、急性心筋梗塞に対するマグネシウム静脈内投与の効果についてのメタアナリシス($n = 7$)を考える。これはメタアナリシスによる結果(有意差あり)と大規模臨床試験の結果(有意差なし)が異なる代表的な例として知られている。

数値実験で比較した4つの手法をもとにオッズ比の95%信頼区間を構成。提案手法の信頼区間はオッズ比1をカバーしてる。

Method	Estimate	95%CI
ECI	0.449	(0.153, 1.094)
REML	0.437	(0.213, 0.895)
DL	0.448	(0.233, 0.861)
LR	0.449	(0.192, 0.903)

2変量モデル (診断法のメタアナリシス)

- ある診断法に対して、 i 番目の研究での感度(sensitivity)、特異度(specificity)をロジット変換した値をそれぞれ y_{Ai}, y_{Bi} とし、それらの分散を s_{Ai}, s_{Bi} とする。このようなデータに対してメタアナリシスを行うためには、以下の2変量モデルを用いる。

$$\begin{pmatrix} y_{Ai} \\ y_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix}, C_i \right), \quad \text{with } C_i = \begin{pmatrix} s_{Ai}^2 & 0 \\ 0 & s_{Bi}^2 \end{pmatrix},$$

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma \right), \quad \text{with } \Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix},$$

興味のあるパラメータ: μ_A, μ_B

一般的に μ_A, μ_B は関連が強いので、個別の信頼区間よりも (μ_A, μ_B) を与える方が有益。

- Reitsma et al. (2005)による $100(1 - \alpha)\%$ の近似的信頼領域

$$\mu_A = \hat{\mu}_A + c_\alpha s_A \cos t, \quad \mu_B = \hat{\mu}_B + c_\alpha s_B \cos(t + \arccos r), \quad t \in [0, 2\pi).$$

ここで s_A, s_B はそれぞれ $\hat{\mu}_A, \hat{\mu}_B$ の標準誤差。

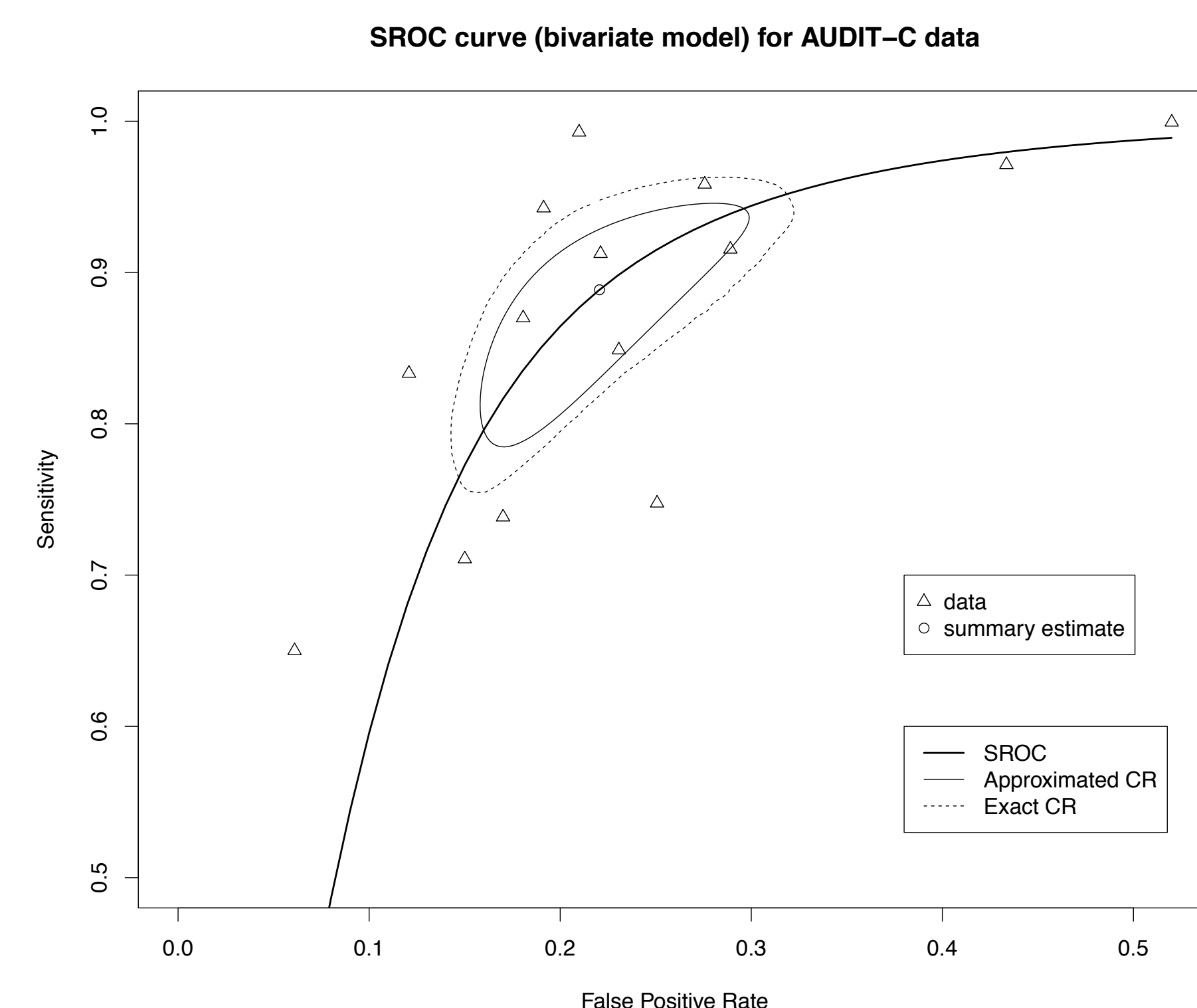
- 単変量のケースと同様に、 $H_0: \mu_A = \mu_{A0}, \mu_B = \mu_{B0}$ の尤度比検定の正確な p 値を計算するアルゴリズムを構成し、その結果をもとに正確な信頼領域を構成する。(詳細は割愛)

- 実データへの適用

アルコール依存症に対するスクリーニングテストの性能に対する $n = 14$ の研究。

感度, 偽陽性率(1-特異度)についての信頼領域をReitsmaの方法(Approximated CR)および提案手法(Exact CR)で求めた。

$n = 14$ でも近似的な信頼領域は、局外母数による不確実性を考慮しきれていないとは言えず、提案手法よりも小さい信頼領域を与えている。



参考文献

- DerSimonian and Laird (1986). *Controlled Clinical Trials*, **7**, 177-188.
- Lindqvist and Taraldsen (2005). *Biometrika*, **92**, 451-464.
- Reitsma et al. (2005). *Journal of Clinical Epidemiology*, **58**, 982-90.
- Teo et al. (1991). *British Medical Journal*, **303**, 1499-1503.