

資料

ベイズモデリングによる合成得点の信頼性係数と
最大信頼性の推定と確信区間の構成

秋 山 隆

Bayesian modeling for composite reliability and maximal reliability.

Takashi AKIYAMA

Abstract

A reliability coefficient in psychometrics is used as an index of consistency. The α coefficient has been widely used as an estimate of reliability coefficient: however, in recent years, there has been an increasing interest in devising other methods of estimating reliability. I have made extensive revisions to enhance clarity and reduce redundancy. In addition to reporting the point estimate of the reliability coefficient, it is also recommended to report the results of interval estimation. Furthermore, psychological research using Bayesian modeling is gradually gaining popularity. In this paper, we introduce a Bayesian model for obtaining the point and interval estimation of maximal reliability and ω coefficient using a statistical analysis environment R and Stan that implements HMC sampling.

Keyword: reliability coefficient, composite reliability, maximal reliability, ω coefficient, Bayesian modeling

抄 録

信頼性係数は心理尺度開発場面で、尺度の安定の度合いを示す指標として利用されている。信頼性係数の代表的な指標として α 係数が広く利用されてきた。近年、 α 係数の再検討が進み、その他の信頼性係数の指標にも関心が高まっている。また、信頼性係数の報告も点推定値のみならず、区間推定を行った結果を報告する事も意識されるようになってきている。更に、ベイズモデリングを利用した心理学研究が増えつつある。本稿では統計解析環境RおよびHMCサンプリングを実装したStanを用いて、ベイズモデリングによって最大信頼性および ω 係数の推定値と確信区間を構成する方法を紹介する。

キーワード：信頼性係数、合成得点の信頼性、最大信頼性、 ω 係数、ベイズモデリング

信頼性

信頼性 (reliability) は、心理・教育測定 (psychological measurement; educational measurement) において、観測対象のテストにおける値の一貫性の高さ、あるいは安定性の程度を表す概念である。信頼性の高低を表現する指標として、信頼性係数 (reliability coefficient) が広く用いられている。

ここでは、表記を簡潔にするため、観測対象の添字を省略する。個人のテストに関する

観測値 x について、個人の真の値（あるいは得点） t との間に、古典的テスト理論（classical test theory）モデル

$$x = t + e \tag{1}$$

を仮定する。誤差 e に関して、その期待値が $E[e]=0$ 、真の得点 t との間に $E[te]=0$ という仮定を置くことで、平均値に関して $\mu_x=E[x]=E[t]=\mu_t$ 、分散に関して $\sigma_x^2=\sigma_t^2+\sigma_e^2$ が導かれる。この下でテストの信頼性係数は

$$\rho = \frac{\sigma_t^2}{\sigma_x^2} \tag{2}$$

と定義される。ただし σ_e^2 は未知であり、 ρ を直接求めることはできないため、推定する必要がある。

α 係数

いま、テストが J 個の項目から構成されており、その合成得点を $y = \sum_{j=1}^J x_j$ とする。合成得点に関する信頼性係数の推定値として現在も広く用いられている α 係数（Cronbach, 1951）は、

$$\alpha = \frac{J}{J-1} \left[1 - \frac{\sum_{j=1}^J \sigma_{x_j}^2}{\sigma_x^2} \right] \tag{3}$$

によって算出される。(3)式は項目が2値で表される場合の信頼性係数の推定公式である、Kuder-Richardsonの公式20（Kuder & Richardson, 1937）を項目得点が3値以上の場合に一般化した指標である。 α 係数は折半信頼性に基づいた指標であり、可能な全てのテストの折半方法で得られる折半信頼性係数の平均値でもある。 α と ρ の間には $\rho \geq \alpha$ という関係があり、(1)式に関して、例えば測定機会2のテスト同士が本質的にタウ等価（essentially tau-equivalent）

$$x_1 = t_1 - \mu_1 + e_1$$

$$x_2 = t_2 - \mu_2 + e_2$$

である場合に等号が成立することが知られている（Novick & Lewis, 1967）。この性質をもってしばしば α は信頼性係数の下限を与えると解釈されてきた。岡田（2015）は α 係数は厳密には推定値の下限ではなく、下界の一つであることを紹介し、 α 係数の意味について詳しい説明を与えている。また、McNeish（2017）も α 係数の仮定と注意点を説明した上で、代替となる各種の信頼性係数の推定方法を紹介している。

ω 係数

本質的にタウ等価測定の設定が満たされない場合、α 係数の推定値にはバイアスが生じるが、その計算の簡便さから、長らく信頼性の推定量として様々な尺度開発研究で用いられてきた。しかし近年では、より仮定の緩い、現実のデータ生成状況に近いと思われる信頼性の推定量である ω 係数への注目が我が国でも高まっている (例えば、清水 (2007)、清水 (2010)、岡田 (2011)、岡田 (2015))。

ω 係数 (McDonald, 1978, 1999) は本質的にタウ等価測定の設定よりも緩い同属 (同族, congeneric) テストの設定の下で、α 係数よりも大きな推定値を与えることが知られており (証明は McDonald, 1999, p.93 に見ることができる)、その利用が広がっている。同属テストの状況では各項目が同一の構成概念から影響を受ける

$$x_j = a_j + \lambda_j f + e_j \quad (4)$$

という因子分析モデルを仮定する。このとき、ω 係数は

$$\omega = \frac{(\sum_{j=1}^J \lambda_j)^2}{(\sum_{j=1}^J \lambda_j)^2 + \sum_{j=1}^J \sigma_{e_j}^2} \quad (5)$$

によって算出される。 λ_j は因子負荷量、 $\sigma_{e_j}^2$ は独自性の分散である。本質的にタウ等価な状況は、(4)式において $\lambda_j = \lambda$ であることを仮定しており、同属テストの状況では項目ごとに因子負荷量や誤差分散が異なることを許容するモデルである。

前述の通り、我が国における心理学の研究場面では信頼性係数の推定値として、より強い仮定を置く α 係数が今も多く用いられており、高本・服部 (2015) は2001年度から2013年度まで、心理学領域の学術誌に掲載された尺度作成に関連した論文189報のうち、168報が α 係数を、3報が ω 係数を信頼性の指標として用いていたことを報告している。岡田 (2011) は α 係数、ω 係数、構造方程式モデルに基づく信頼性係数の推定方法について、シミュレーションによって各種推定値のバイアスの程度を比較し、真のモデル構造が自明ではない場合は ω 係数 (岡田 (2011) では ω_t と表記) の利用を推奨している。また、Raykov (2019) は、 $\rho - \alpha = \epsilon$ とし、誤差間相関を伴わない同属テストの状況下では、サンプルサイズ n の下で α 係数の推定量 $\hat{\alpha}_n$ が、

$$\hat{\alpha}_n \xrightarrow{a.s.} \alpha = \rho - \epsilon < \rho, \quad (n \rightarrow \infty \text{ のとき})$$

へと強収束 (strong convergence) することを指摘している。ここで、 $\xrightarrow{a.s.}$ はほとんど確実に (converge almost surely, つまり確率 1 で) 収束することを表している。

重み付けた場合の信頼性係数の推定

w_j が既知の場合

x_j を重み付けによって再尺度化した $y_j=w_j x_j$ に関して、 $\mathbf{w}=(w_1, \dots, w_J)'$ とすると、(2)式に対応する、重み付き合成得点 $y = \sum_{j=1}^J w_j x_j$ の信頼性は

$$\rho_y = \frac{\sigma_{t_y}^2}{\sigma_y^2} = 1 - \frac{\mathbf{w}'\Psi_e \mathbf{w}}{\mathbf{w}'\Sigma_y \mathbf{w}}$$

と定義される (Mosier, 1943; Bentler, 2007)。合成得点を単純和 ($w_j=1$) とすると信頼性係数は

$$\rho = \frac{\sigma_t^2}{\sigma_x^2} = 1 - \frac{\mathbf{1}'\Psi_e \mathbf{1}}{\mathbf{1}'\Sigma \mathbf{1}}$$

と再表現される。 ρ_j を要素得点 x_j の信頼性とし、 $\sigma_{e_j}^2 = \sigma_{x_j}^2(1 - \rho_j)$ とすれば

$$\rho = 1 - \frac{1}{\sigma_x^2} \left[\sum_{j=1}^J w_j^2 \sigma_{x_j}^2 (1 - \rho_j) \right]$$

である。なお、項目 j の重みを w_j とした重み付き合成得点 $\sum_{j=1}^J w_j x_j$ の信頼性に関する α 係数は

$$\alpha_c = \frac{J}{J-1} \left[1 - \frac{\sum_{j=1}^J w_j^2 \sigma_{x_j}^2}{\sigma_x^2} \right]$$

によって算出される。

w_j が未知の場合

w_j が未知の場合、 w_j の定め方として、素点の和 ($w_j=1$)、標準得点の合計、 α_c を最大にする方法、合成得点の信頼性を最大にする方法などがある (池田, 1973, pp.247-255)。これらのうち、信頼性を最大とする重みは $\mathbf{C}^{-1}\tilde{\mathbf{C}}$ の最大固有値 d に対応した固有ベクトルを固有値・固有ベクトル方程式

$$(\tilde{\mathbf{C}} - d\mathbf{C})\mathbf{w} = (\mathbf{C}^{-1}\tilde{\mathbf{C}} - d\mathbf{I})\mathbf{w} = \mathbf{0}$$

を解くことで求められる。 \mathbf{C} は要素得点 x_j についての分散共分散行列、 $\tilde{\mathbf{C}}$ は \mathbf{C} の対角成分を各要素の信頼性係数に置き換えた行列である (Green, 1950; 池田, 1973)。いま、 \mathbf{R} を \mathbf{C} に

対応した相関行列とし、 $\tilde{\mathbf{R}}$ を $\tilde{\mathbf{C}}$ に対応した、その対角成分に各要素の信頼性係数を配した相関行列とする。このとき

$$\rho_{max} = \frac{\mathbf{w}'\tilde{\mathbf{R}}\mathbf{w}}{\mathbf{w}'\mathbf{R}\mathbf{w}}$$

は最大信頼性 (maximal reliability) と呼ばれる (Li, Rosenthal, & Rubin, 1996; Li, 1997)。最大信頼性もまた、 ω 係数と並び、近年関心を集めている (例えば Raykov, Marcoulides, & Gabler (2017))。実際場面では $\tilde{\mathbf{R}}$ の対角要素には信頼性の推定値が配されることになり、Thissen & Wainer (2001) はこれに α 係数を用いた計算例を示している。

最大の信頼性を達する、最適な重みを見出そうとする手法についての研究は、Mosier (1943) や Green (1950) にまで遡ることができる。Haertel (2006) は Lord & Novick (1968) に基づいて最大信頼性と重みを得る方法を紹介し、Li (1997)、Li et al. (1996) を信号雑音比の観点から再母数化したものとして紹介している。また、Bentler (2007) は Green (1950) や Bentler (1968) に基づいて最大信頼性を紹介したのち、「再発見された (rediscovered)」という言葉を用いて Liらの文献を紹介している。Thissen & Wainer (2001) は、最大の信頼性を達成する重みを得る方法に関する研究は1940年代から行われてきたにも関わらず、かつてはその利用が広まらなかった一因に、固有値を求める計算の煩雑さを指摘している。しかし現在では、この点に関してはPCの性能向上と、安価に利用可能な、行列演算を行えるソフトウェアの発展によって解消されたものと言えるだろう。また、次に紹介するように、確認的因子分析モデルの観点から最大信頼性を与える重みを推定する方法を利用すれば、構造方程式モデリングのソフトウェアを用いて計算することも可能である。

構造方程式モデリングに基づいた最大信頼性の表現

Raykov (2004) は構造方程式モデルを用いて、確認的因子分析モデルを仮定し、因子負荷量の推定値を用いて、合成得点の重みと最大信頼性を推定する方法を提案した。また、構造方程式モデルを記述・分析するためのプログラムである LISREL のコードを提供している。また、Raykov (2006) は最大信頼性について区間推定を行う方法を論じている。重み付き合成得点 $y = \sum_{j=1}^J w_j x_j$ に関して、最大信頼性は信号雑音比の観点から

$$\rho_{max} = \frac{\sum_{i=1}^J (\rho_j / (1 - \rho_j))}{\delta + \sum_{j=1}^J (\rho_j / (1 - \rho_j))}, \quad w_j = \frac{\rho_j}{\lambda_j (1 - \rho_j)}$$

と表現される (Li, 1997; Raykov, 2004)。 δ は母数として定義されているが、同属テストの場合 $\delta=1$ となる (Li, 1997)。いま、合成変数の要素 x に関して、(4)式の因子分析モデル $x_j = a_j + \lambda_j f + e_j$ を仮定する。因子の分散について $\sigma_f^2 = 1$ とし、 $\sigma_e^2 = \lambda_j^2$ であることを考慮すると ρ_j は

$$\rho_j = \lambda_j^2 \frac{\sigma_f^2}{\sigma_{x_j}^2} = \frac{\lambda_j^2}{\lambda_j^2 + \sigma_{e_j}^2}$$

と表される。上式を w_j の ρ_j に代入することで、

$$w_j = \frac{\rho_j}{\lambda_j(1 - \rho_j)} = \frac{\lambda_j^2 / (\lambda_j^2 + \sigma_{e_j}^2)}{\lambda_j \sigma_{e_j}^2 / (\lambda_j^2 + \sigma_{e_j}^2)} = \frac{\lambda_j}{\sigma_{e_j}^2} \quad (6)$$

を得る。また、真の値に関して w_j で重み付けた結合和を $T = \sum_{j=1}^J w_j t_j$ とするとき、 y の信頼性係数はその定義から、

$$\rho_y = \frac{\sigma_T^2}{\sigma_y^2} = \text{Cor}(T, y)^2 = \text{Cor}\left(\sum_{j=1}^J w_j a_j + \left(\sum_{j=1}^J w_j \lambda_j\right) f, y\right)^2 = \text{Cor}(f, y)^2 \quad (7)$$

と表される。つまり、確認的因子分析モデルと、合成得点 $y = \sum_{j=1}^J w_j x_j$ を仮定し、 $w_j = \lambda_j / \sigma_{e_j}^2$ と制約することで、最大信頼性を得る重みを推定することができる。また、 $\rho_{\max} \geq \omega$ (Raykov, Gabler & Dimitrov, 2015, Appendix B)、 $\omega \geq \alpha$ (McDonald, 1999, p.93) から、 $\alpha \leq \omega \leq \rho_{\max}$ である。つまり、最大信頼性も本質的にタウ等価の仮定が満たされない場合、 $\rho_{\max} \geq \alpha$ となる信頼性の推定値を与える。以降では、構造方程式モデルの下位モデルとしての因子分析モデルによる表現に基づき、ベイズモデリングによって ω 係数と最大信頼性を得る方法を紹介する。

ベイズモデリングと事後分布

観測対象 $i (=1 \dots N)$ の項目 $j (=1 \dots J)$ に対する反応を縦に並べた $J \times 1$ 反応パターンベクトル $\mathbf{x}_i = (x_{i1}, \dots, x_{ij})'$ が因子分析モデル $\mathbf{x}_i = \mathbf{a} + \mathbf{\Lambda} f_i + \mathbf{e}_i$ によって得られるものとする。 \mathbf{a} は $J \times 1$ の切片ベクトル $(a_1, \dots, a_j)'$ 、 $\mathbf{\Lambda}$ は $J \times 1$ の因子負荷量を並べたベクトル $(\lambda_1, \dots, \lambda_j)'$ 、 f_i は観測対象 i の因子スコア、 \mathbf{e}_i は観測対象 i の独自因子スコアベクトル $(e_{i1}, \dots, e_{ij})'$ である。 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N)'$ とするとき、反応パタンの尤度は

$$L(\mathbf{X} | \mathbf{a}, \mathbf{\Lambda}, f_i, \mathbf{e}_i) = \prod_{i=1}^N \prod_{j=1}^J g(x_{ij} | a_j, \lambda_j, f_i, e_{ij})$$

と表される。ここで、 $E[f_i] = 0, E[f_i^2] = 1, E[\mathbf{e}_i] = \mathbf{0}, E[\mathbf{e}_i \mathbf{e}_i'] = \mathbf{\Sigma}_e, E[f_i \mathbf{e}_i] = \mathbf{0}$ とし、 \mathbf{x}_i が J 次

元の変量正規分布 $\mathbf{x}_i \sim MN(\mathbf{a} + \Lambda \mathbf{f}_i, \Sigma_e)$ に従うものとする。 Σ_e は対角要素に独自因子の分散 ψ_j^2 を配した対角行列である。いま、 \mathbf{a} 、 Λ 、 Σ_e 間に

$$g(\mathbf{a}, \Lambda, \Sigma_e) = g_a(\mathbf{a})g_\Lambda(\Lambda)g_e(\Sigma_e)$$

が成り立つとする。事前分布はそれぞれ、

$$g_a(\mathbf{a}) = \prod_{j=1}^J g(a_j | \mu_a, \sigma_a^2), \quad g_\Lambda(\Lambda | \mu_\lambda, \sigma_\lambda^2) = \prod_{j=1}^J g(\lambda_j | \mu_\lambda, \sigma_\lambda^2), \quad g_e(\Sigma_e) = \prod_{j=1}^J g(\psi_j^2)$$

とすると、事後分布は

$$\begin{aligned} g(\mathbf{a}, \Lambda, \mathbf{f}_i, \mathbf{e}_i | \mathbf{X}) &\propto L(\mathbf{X} | \mathbf{a}, \Lambda, \mathbf{f}_i, \mathbf{e}_i) g(\mathbf{a}, \Lambda, \Sigma_e) \\ &= \prod_{i=1}^N \left[\prod_{j=1}^J g(x_{ij} | a_j, \lambda_j, \mathbf{f}_i, \mathbf{e}_{ij}) g(a_j | \mu_a, \sigma_a^2) g(\lambda_j | \mu_\lambda, \sigma_\lambda^2) g(\psi_j^2) \right] \\ &\quad \times g(\mu_a) g(\sigma_a^2) g(\mu_\lambda) g(\sigma_\lambda^2) \end{aligned}$$

と導かれる。

生成量

T 個の事後分布からのサンプルを用い、(6)式に基づいて、生成量として合成得点 $\mathbf{y}_i^{(t)} = \mathbf{w}^{(t)'} \mathbf{x}_i^{(t)}$ を定義する。上付きの (t) は連鎖内における t 番目のサンプルを表す。このとき、因子負荷量および独自性の分散の逆数を用いることで、最大信頼性を得る重み $w_j^{(t)} = \lambda_j^{(t)2} / \psi_j^{(t)2}$ に関する生成量および事後分布の近似を得る。同様に、最大信頼性を(7)式に基づき、

$$\rho_y^{(t)} = \text{Cor}(\mathbf{w}^{(t)'} \mathbf{a}^{(t)} + (\mathbf{w}^{(t)'} \boldsymbol{\lambda}^{(t)}) \mathbf{f}^{(t)}, \mathbf{y}^{(t)})^2$$

のように生成量として定義することで、その事後分布 $g(\rho_y)$ の近似を得ることができる。応用場面において、信頼性係数の指標はしばしばその値のみが報告されてきた。しかしながら、それは標本反応パターンデータから推定された点推定値に過ぎず、誤差を伴うため、区間推定の結果を報告することも重要である。高本・服部 (2015) は2000年代以降に、信頼性係数を区間推定することの重要性を指摘した複数の文献を紹介している。また、Raykov (2006) は構造方程式モデリングの観点から表現した最大信頼性について、標準誤差を用いて信頼区間を構成する方法を示している。Trinchera, Marie, & Marcoulides (2018) は α 係数の信頼区間の求め方の各手法を概観し、漸近分布に基づいた信頼区間の構成方法を提案している。

一方で、ベイズモデリングによって、点推定値のみならず、推定値の区間を事後分布から直接構成し、同時に考察の対象とすることができる。例えば、生成量の事後分布の近似として得た $g(\rho_j)$ について、点推定値として EAP (expected a posteriori, 事後期待値) 推定値や MAP (maximum a posteriori, 事後確率最大値) 推定値を利用し、区間推定のためには事後分布の95%確信区間や HDI (highest density interval; 最高密度区間) を構成する。

実際の応用場面では、事後分布からのサンプルを得るために、ハミルトニアンモンテカルロ (Hamiltonian Monte Carlo; Hybrid Monte Carlo, HMC; Duane, Kennedy, Pendleton, & Roweth, 1987) 法を利用することができる。HMC 法はベイズモデリング用のプログラミング言語と HMC サンプリングを実装したソフトウェア Stan (Stan Development Team, 2019) と、統計解析環境 R (R Core Team, 2019) および Stan の R 用インターフェースを提供するパッケージ RStan (Stan Development Team, 2018) を用いて実行可能であり、本稿付録において、事後分布からの HMC サンプルを要約することで最大信頼性の点推定値と確信区間を得るコードを提供している。なお、本稿付録では Stan コード内の generated quantities ブロックではなく、R において生成量の要約を行なっている。

前述のベイズモデリング (および付録の Stan コード) はあくまで同属テストモデルを表現しているため、(5)式に基づき、因子負荷量 λ_j と独自因子の標準偏差 ψ_j の HMC サンプルを用いて ω 係数の生成量

$$\omega^{(t)} = \frac{(\sum_{j=1}^J \lambda_j^{(t)})^2}{(\sum_{j=1}^J \lambda_j^{(t)})^2 + (\sum_{j=1}^J \psi_j^{(t)2})}$$

を定義することも可能である。本稿付録にて ω 係数生成のための R コードを提供する。

シミュレーション

R 用パッケージ semTools (ver.0.5.1, Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2018) において提供されている maximalRelia 関数を用いて推定した最大信頼性と、付録に示す Stan コード (Stan, ver.2.19) によって得た、事後分布からのサンプルを用いて構成した最大信頼性の EAP 推定値との乖離の程度をシミュレーションを通じて確認した。maximalRelia 関数は Li (1997) の方法に基づいて最大信頼性と最適尺度の重みを推定する。シミュレーションデータ 本稿では $N=300$, 項目数を 6 とし、岡田 (2011) の設定を参考にし、表 1 の状況設定に基づいてシミュレーションのための 100 セット分の反応パターンデータを発生させた。なお岡田 (2011) は比較可能性のために、Yang & Green (2010) の設

表1 シミュレーションデータ発生状況

モデル	因子負荷量	真の信頼性
本質的にタウ等価モデル	全て0.3	0.372
	全て0.8	0.914
同属モデル	0.3から0.8	0.731

定を踏襲している。なお、表1の第3列は真の最大信頼性を表しているわけではないことに注意してほしい。

事前分布および Stan の設定 複数連鎖間で因子負荷量の符号が反転する可能性があるため、シミュレーションにあたり、因子負荷量の範囲を正の値に制約することとした。また、各連鎖の初期値として $\Lambda_j=0$ を与えている。シミュレーション回数はサンプリングに必要な時間とPCの性能を考慮し、100とした。HMCの連鎖数を5、各連鎖の更新回数を2000、うち最初の1000回をウォームアップ期間として破棄し、残りの5000個のサンプルを用いて事後分布を構成することとした。

$g(a_j|\mu_a, \sigma_a)$ として、平均 $\mu_a=0$ 、標準偏差 $\sigma_a=100$ の正規分布を仮定した。 $g(\lambda_j|\mu_\lambda, \sigma_\lambda)$ には平均 μ_λ 、標準偏差 σ_λ の正規分布を仮定した。ただし $\mu_\lambda=0$ とし、 σ_λ の超事前分布として Stan による無情報的一様分布を仮定した。 $g(\psi_j)$ も同様に Stan による無情報的一様分布を仮定した。また、因子スコアの事前分布として $N(0,1)$ を仮定した。

結果

全シミュレーションにおいて、収束判定指標である \hat{R} (Gelman & Rubin, 1992; Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2014, p.285) は1.1を下回っていたため、連鎖は収束しているものと見なし、得られたサンプルを用いて最大信頼性を算出した。 $s(=1, \dots, S)$ 回目のシミュレーションにおけるベイズモデリングに基づく最大信頼性の点推定値を $\rho_{(mr.eap)s}$ とし、R関数 maximalRelia によって算出した点推定値を $\rho_{(MR)s}$ とする。シミュレーションにおける乖離の程度の平均

$$S^{-1} \sum_{s=1}^{100} [(\rho_{(mr.eap)s} - \rho_{(MR)s}) / \rho_{(MR)s}]$$

は、本質的にタウ等価モデルの下で -0.000108 、同属テストモデルの下で -0.000252 であり、点推定値ではパッケージによる推定結果と近い値を得られる。また、 α 係数との差を比較したところ、100回の全てにおいて $\rho_{mr.eap} > \alpha$ であることが確認された。図1に最大信

頼性のEAP推定値と、maximalRelia関数による推定値の散布図を示した。図1左は同属モデルの状況を、右は本質的にタウ等価モデルの状況を表している。推定値間の相関係数は同属モデルの場合で0.999、本質的にタウ等価モデルの場合で1.000であった。

図2はシミュレーションで発生させたデータセットに対して、最大信頼性のEAP推定値と ω 係数のEAP推定値をプロットした散布図である。図2左は同属テストの状況を、右は本質的にタウ等価モデルの状況を表している。推定値間の相関係数は同属テストの場合で0.742、本質的にタウ等価モデルの場合で0.995であった。本質的にタウ等価な場合には同属モデルの場合と比較して、推定値の相違は小さい。ただし、どちらの状況においても、100回の推定において、最大信頼性が ω 係数よりも大きくなっている様子が観察された。最

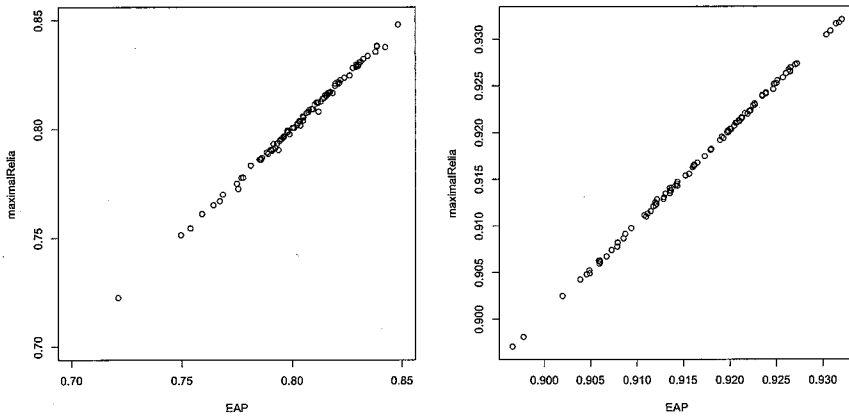


図1 EAP推定値とmaximalRelia関数による推定値の散布図（左：同属、右：本質的にタウ等価）

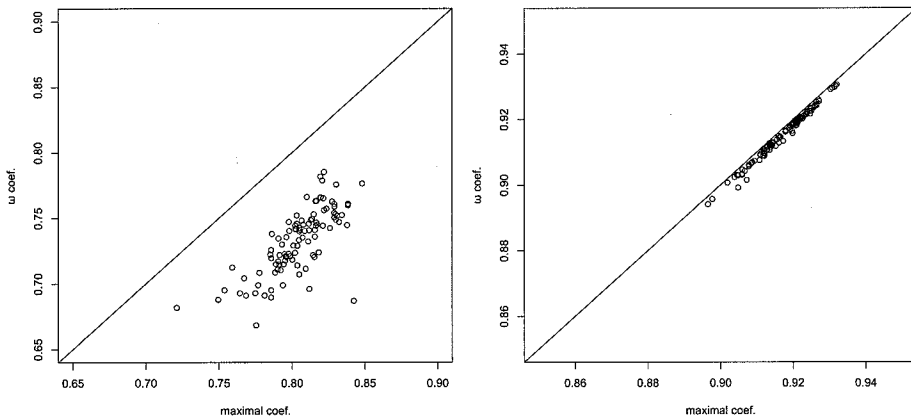


図2 最大信頼性と ω 係数のEAP推定値の散布図（左：同属、右：本質的にタウ等価）

大信頼性の EAP 推定値の、 ω 係数に対する差異の程度

$$\frac{\rho_{maxeap} - \omega_{eap}}{\omega_{eap}}$$

の相加平均は同属モデルの場合で平均的に約9.9% (0.099) 大きく、本質的にタウ等価モデルの場合で平均的に約0.2% (0.002) 大きかった。

最大信頼性は α 係数と同様に長い歴史を有しているが、 α 係数に代わる信頼性の指標として再度注目が集まったのは近年であり、その性質に関して検討した研究は未だ十分とは言えない。Aguirre-Urreta, Rönkkö, & McIntosh (2019) は有限標本下における最大信頼性と合成得点の信頼性 (ω 係数) の推定値の振る舞いについて、サンプルサイズ、変数の数、因子負荷量の値の条件水準を組み合わせるシミュレーション研究を行なっている。その結果、最大信頼性は母集団における最大信頼性の値に対して、有限標本下では推定値に正のバイアスが生じ、サンプルサイズを増加させていくに従い、値が母集団の値へと近づいていく様子が報告されている。また、バイアスの程度は合成得点の信頼性と比較して、相対的に大きいことも示されている。Aguirre-Urreta et al. (2019) は、サンプルサイズ条件以外の条件水準変化時の総合的な結果から、実際の応用場面において、信頼性係数をモデルの質を評価する指標として使用する場合には合成得点の信頼性を利用することを推奨している。最大信頼性に関して、バイアス修正の試みもなされており (Penev & Raykov, 2010)、今後の更なる研究の発展が望まれる。

確信区間の構成

例として、付録 2 で定義した `sumres` 関数を用いて、シミュレーションデータのうちの 1 つに関して、最大信頼性の生成量のサンプルを要約した結果を以下に示す。ここで `rhomax` は最大信頼性の生成量についての、長さ 5000 のベクトルオブジェクトである。 ω 係数に対しても、同じ関数を利用して区間推定を行うことができる。

```
> round(sumres(rhomax), 3)
      EAP post. sd  2.5%   50%  97.5% HDIlow HDIup
0.814  0.020  0.774  0.815  0.851  0.775  0.852
```

HDIlow と HDIup はそれぞれ 95% HDI の下限と上限を表している。この例では、95% 確信区間と HDI の間に大きな違いは見られなかった。MCMC サンプルを用いた信頼性係数

の推定では、区間推定も事後分布の要約を通じて行うことができる。

また、任意の信頼性係数の推定量についての、 t 番目の生成量サンプルを $\hat{\rho}^{(t)}$ とするとき、あるテストの信頼性係数の推定値が閾値 c より大きくなる確率（豊田, 2017）

$$u_{\hat{\rho} > c}^{(t)} = h(\hat{\rho}^{(t)}) = \begin{cases} 1 & \hat{\rho}^{(t)} > c \\ 0 & \text{otherwise} \end{cases}$$

$$p(\hat{\rho} > c) = T^{-1} \sum_{t=1}^T u^{(t)}$$

を評価することもできる。例えば、上述のシミュレーションデータの場合、閾値を0.75、0.80、0.85とした場合、それぞれの閾値に対して

```
> mean(ifelse(rhmax > 0.75, 1, 0))
[1] 0.9986
> mean(ifelse(rhmax > 0.80, 1, 0))
[1] 0.7606
> mean(ifelse(rhmax > 0.85, 1, 0))
[1] 0.0264
```

となり、0.75より大きくなる確率は約99.9%、0.85よりも大きくなる確率は約2.6%であると解釈できる。

引用文献

- Aguirre-Urreta, M. I., Rönkkö, M., & McIntosh, C. N. (2019). A cautionary note on the finite sample behavior of maximal reliability. *Psychological Methods*, *24* (2), 236-252. <https://doi.org/10.1037/met0000176>
- Bentler, P. M. (1968). Alpha-maximized factor analysis (alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika*, *33* (3), 335-345.
- Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In Lee, S.-Y. (Eds.) *Handbook of latent variable and related models*. North-Holland: Amsterdam, The Netherlands. 1-19.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16* (3), 297-334.
- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters*, *B* (195), 216-222.
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7* (4), 457-472. doi:10.1214/ss/1177011136.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data*

- analysis, Third edition*. CRC Press: Boca Raton, FL.
- Green, B. F. (1950). A note on the calculation of weights for maximum battery reliability. *Psychometrika*, **15** (1), 57-61.
- Haertel, E. H. (2006). Reliability. In Brennan, R. L. (Eds.) *Educational measurement (Fourth edition)*. American Council on Education and Praeger: Westport, CT. 65-110.
- 池田央 (1973). 心理学研究法 8 — テスト II — 東京大学出版会
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-1. Retrieved from <https://CRAN.R-project.org/package=semTools>
- Kuder, G. F. & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, **2** (3), 151-160.
- Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, **1** (1), 98-107.
- Li, H. (1997). A unifying expression for the maximal reliability of a linear composite. *Psychometrika*, **62** (2), 245-249.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McDonald, R. P. (1978). Generalizability in factorable domains: "Domain validity and generalizability". *Educational and Psychological Measurement*, **38** (1), 75-79.
- McDonald, R. P. (1999). *Test theory: a unified treatment*. Hillsdale: Erlbaum.
- McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, **23** (3), 412-433.
- Mosier, C. L. (1943). On the reliability of a weighted composite. *Psychometrika*, **8** (3), 161-168.
- Novick, M. R. & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, **32** (1), 1-13. <https://doi.org/10.1007/BF02289400>
- 岡田謙介 (2011). クロンバックの α に代わる信頼性の推定法について — 構造方程式モデリングによる方法・McDonald の ω の比較 — 日本テスト学会誌 **7** (1), 37-50.
- 岡田謙介 (2015). 心理学と心理測定における信頼性について — Cronbach の α 係数とは何なのか、何でないのか — 教育心理学年報 **54**, 71-83.
- Penev, S. & Raykov, T. (2010). A method of bias correction for maximal reliability with dichotomous measures. *British Journal of Mathematical and Statistical Psychology*, **63** (1), 163-175.
- Raykov, T. (2004). Estimation of maximal reliability: A note on a covariance structure modeling approach. *British Journal of Mathematical and Statistical Psychology*, **57**, 21-27.
- Raykov, T. (2006). Interval estimation of optimal scores from multiple-component measuring instruments via SEM. *Structural Equation Modeling*, **13** (2), 252-263.
- Raykov, T., Gabler, S., & Dimitrov, D. M. (2015). Maximal reliability and composite reliability: Examining their difference for multicomponent measuring instruments using latent variable modeling. *Structural Equation Modeling*, **23** (3), 384-391. <https://doi.org/10.1080/10705511.2014.966369>
- Raykov, T., Marcoulides, G. A., & Gabler, S. (2017). Improved estimation of maximal reliability for unidimensional multicomponent measuring instruments in repeated measure studies. *Structural Equation Modeling*, **24** (5), 755-767.

- Raykov, T. (2019). Strong convergence of coefficient alpha estimator for reliability of multiple-component measuring instruments. *Structural Equation Modeling*, 26 (3), 430-436.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- 清水和秋 (2007). a はやめて ω にしよう — 因子分析で構成した尺度の共通性と信頼 — 日本心理学会第71回大会発表論文集
- 清水和秋 (2010). 項目因子分析で構成した尺度の因子パターン, 共通性, 信頼性そして因子的事実性 関西大学心理学研究 1, 9-24.
- Stan Development Team. 2019. Stan Modeling Language Users Guide and Reference Manual, Version 2.19.0. Retrieved from <http://mc-stan.org>
- Stan Development Team. 2018. RStan: the R interface to Stan. R package version 2.18.2. Retrieved from <http://mc-stan.org>
- 高本真寛・服部環 (2015). 国内の心理尺度作成論文における信頼性係数の利用動向 心理学評論 58 (2), 220-235.
- Thissen, D. & Wainer, H. (2001). *Test scoring*. Lawrence Erlbaum: Mahwah, NJ.
- 豊田秀樹 (2017). p 値を使って学術論文を書くのは止めよう 心理学評論 60 (4), 379-390.
- Trincherá, L., Marie, N., & Marcoulides, G. A. (2018). A distribution free interval estimate for coefficient alpha. *Structural Equation Modeling*, 25 (6), 876-887. <https://doi.org/10.1080/10705511.2018.1431544>
- Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, 17 (1), 66-81.
- Yves R. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>.

—2019.6.29受稿—

付録 1

最大信頼性と ω 係数推定のために用いた Stan コードを以下に示す。

```

data {
  int N; //回答者数
  int J; //項目数
  matrix[N, J] Y; //データ行列
}
parameters {
  vector[J] a; //切片
  vector[J] lam; //因子負荷量
  real f[N]; //scores
  real<lower=0> sig_lam; //負荷量 SD
  vector<lower=0>[J] psi; //独自因子 SD
}
transformed parameters {
  matrix[N, J] mu;
  vector[J] w; //重み
  for (i in 1:N) {
    for (j in 1:J) {
      mu[i, j] = a[j] + lam[j] * f[i];
    }
  }
  for (j in 1:J)
    w[j] = lam[j] * (1/pow(psi[j], 2));
}
model {
  for (j in 1:J) {
    a[j] ~ normal(0, 100);
    lam[j] ~ normal(0, sig_lam);
  }
  for (i in 1:N)
    f[i] ~ normal(0, 1);
  for (i in 1:N) {
    for (j in 1:J) {
      Y[i, j] ~ normal(mu[i, j], psi[j]);
    }
  }
}

```

付録 2

Stan によるサンプリング後、下記 R コードを用いて最大信頼性の EAP、95% 確信区間、95%HDI を構成する。

```

Y #Y には行に個体、列に変数を配したデータを付値する。
# 例えば Y <- read.table("hoge.hoge.txt", header=F, sep="\t")
J <- ncol(Y) #観測変数の数
dat2 <- list(N=nrow(Y), J=ncol(Y), Y=Y)
iter <- 4000 #サンプリング数
warmup <- 2000 #ウォームアップ期間
nchain <- 5 #連鎖の構成数
chainsize <- (iter-warmup)*nchain
initl <- "random" #初期値をランダムとする
stanfit <- stan(file="maximalrel.stan", data=dat2, chains=nchain, iter=iter, warmup=warmup, init=initl, seed=123, thin=1)
summary(stanfit, pars=c("a", "lam", "sig_lam"))
traceplot(stanfit, pars=c("a", "lam", "sig_lam"))
# 最適尺度重み
y.weightedlist <- as.list(numeric(chainsize))
Y <- as.matrix(dat2$Y)
w <- extract(stanfit)$w
colMeans(w)
head(w[,1] * Y)
for(ite in 1:chainsize){
  y.weightedlist[[ite]] <- t(w[ite,] * t(Y))
}
ycolmlist <- lapply(y.weightedlist, rowSums)

# 最大信頼性の生成量
a <- extract(stanfit)$a
lam <- extract(stanfit)$lam
f <- extract(stanfit)$f
t <- colSums(w*a) + colSums((w*lam)*f)
rhomax <- numeric(chainsize)
for(ite in 1:chainsize){
  rhomax[ite] <- cor(t[ite,], ycolmlist[[ite]])^2 # 最大信頼性 EAP
}

# 生成量要約のための関数定義
#install.packages("HDInterval") #パッケージ HDInterval を未インストールの場合のみ実行
sumres <- function(xvec, cm=0.95){
  result <- c(
    mean(xvec), sd(xvec), quantile(xvec, prob=c(0.025, 0.50, 0.975)),
    HDInterval::hdi(rhomax, credMass=cm)
  )
  names(result) <- c("EAP", "post.sd", "2.5%", "50%", "97.5%", "HDI low", "HDI up")
  return(result)
}
round(sumres(rhomax), 3)

```


付録 3

maximalRelia 関数で最大信頼性を推定するために用いた、lavaan による分析モデルを下記に付す。

```
analysisModel <- "  
f1 =~ y1 + y2 + y3 + y4 + y5 + y6;  
f1 ~ 1*f1;  
y1 ~ y1; y2 ~ y2; y3 ~ y3;  
y4 ~ y4; y5 ~ y5; y6 ~ y6;  
y1~1; y2~1; y3~1; y4~1; y5~1; y6~1;  
"
```

付録 4

付録 2 で作成した stanfit オブジェクトを用いて、 ω 係数の生成量を作成するための R コードを下記に示す。

```
lam <- extract(stanfit)$lam  
psi <- extract(stanfit)$psi #標準偏差  
omegapost <- (rowSums(lam)^2 / (rowSums(lam)^2 + rowSums(psi^2)))#生成量  
# 付録 2 で定義した sumres を利用した要約  
sumres(omegapost)
```

