

2022.09

テスト尺度信頼性再考

方向性と不確かさの観点から

○岡本安晴
(日本女子大学)

問題と提案

信頼性は、測定対象概念の真値と測定値との関係を表す概念であると考えられる。信頼性係数は、信頼性を表す指標として誤差モデルに基づいて議論されている（Levy & Mislevy, 2016; Raykov & Marcoulides, 2011）。しかし、誤差モデルは潜在クラスなどの質的値には適用できない。また、測定対象概念値に対して測定値が確率的に変動するとき、変動する測定値はいずれも測定対象概念値（以後、概念値と呼ぶ）に対するものであり、その意味でいずれも対等である。それら測定値の平均値は統計学的に定義されるが、平均値とか測定の特定の値を真値とする根拠を示す必要がある。概念値と測定値との関係を考えるときには、誤差モデルは不適當であると言える。

信頼性を誤差モデルに基づいて考えるのではなく、概念値あるいは測定値の不確かさの観点から検討を試み、テスト／尺度による不確かさ（uncertainty）の減少（情報量：Cover & Thomas, 2006）として信頼性を考える。

不確かさは確率分布で表される。概念値 θ と測定値 X の同時確率 $P(\theta, X)$ を、 θ の確率分布 $P(\theta)$ と、 θ をパラメータとする測定値 X の確率分布 $P(X; \theta)$ に対して、式

$$P(\theta, X) = P(\theta)P(X; \theta)$$

で与える。このとき、

$$P(X|\theta) = P(X; \theta)$$

が成り立つ。

確率分布に対して、その不確かさ U は、量的値の確率分布の場合は分散、質的値の確率分布の場合はエントロピーで表すことができる。

概念値 θ と測定値 X の関係は、
概念値 θ から見た測定値 X の確率分布 $P(X|\theta)$ の表す不確かさ

$$U(X|\theta)$$

と、測定値 X から見た概念値 θ の確率分布 $P(\theta|X)$ の不確かさ

$$U(\theta|X)$$

を区別することができる。

信頼性係数を、測定値 X の確率分布 $P(X)$ の不確かさ $U(X)$ の変化量

$$U(X) - U(X|\theta)$$

および概念値 θ の確率分布 $P(\theta)$ の不確かさ $U(\theta)$ の変化量

$$U(\theta) - U(\theta|X)$$

を $U(X)$ および $U(\theta)$ に対する相対量で表すと次式のようにになる。

$$R_c(\theta) = \frac{U(X) - U(X|\theta)}{U(X)}, \quad R_m(X) = \frac{U(\theta) - U(\theta|X)}{U(\theta)}$$

$R_c(\theta)$ は、測定値 X の不確かさ $U(X)$ に対して、概念値 θ に対しての測定値 X の不確かさ $U(X|\theta)$ がどの程度の減少であるのかを $U(X)$ に対する相対値で表すものである。

$R_m(X)$ は、概念値 θ の不確かさ $U(\theta)$ に対して、測定値 X に対しての概念値 θ の不確かさ $U(\theta|X)$ がどの程度の減少であるのかを $U(\theta)$ に対する相対値で表すものである。

これらの信頼性係数の、古典テスト理論、順序カテゴリ尺度、項目反応理論および潜在クラス尺度において算出例を以下に報告する。

信頼性係数の算出

古典テスト理論

古典テスト理論の場合、同時確率は

$$P(\theta, X) = P(\theta)P(X; \theta) = N(\theta; 0, 1)N(X; a + b\theta, \sigma^2)$$

で与えられる。不確かさを分散で与えると、

$$R_c(\theta) = \frac{U(X) - U(X|\theta)}{U(X)} = \frac{b^2}{b^2 + \sigma^2} = \omega$$

また、

$$P(\theta|X) = N\left(\theta; \frac{(X - a)b}{\sigma^2 + b^2}, \frac{\sigma^2}{\sigma^2 + b^2}\right)$$

より、

$$R_m(X) = \frac{U(\theta) - U(\theta|X)}{U(\theta)} = \frac{1 - \frac{\sigma^2}{\sigma^2 + b^2}}{1} = \frac{b^2}{b^2 + \sigma^2} = \omega$$

ここで、 ω は古典テスト理論における信頼性係数である。

順序カテゴリ尺度

順序カテゴリ尺度のモデルとしてサーストンのカテゴリ判断のモデルを設定する (e.g., Okamoto, 2013)。

測定対象概念値を θ 、M個のテスト項目を $X_i; i = 1, \dots, M$ 、尺度値 X を項目和

$$X = \sum_{i=1}^M X_i$$

で表す。

測定対象概念値 θ と測定値 (尺度値) X の同時確率 $P(\theta, X)$ を求める。簡単のため各項目 X_i が2値の場合を考える。

項目 X_i に対する潜在変数を

$$Z_i = a_i + b_i\theta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

とおき、

$$X_i = \begin{cases} 1, & Z_i > 0 \\ 0, & Z_i \leq 0 \end{cases}$$

とする。

Okamoto, Y. (2013). A Direct Bayesian Estimation of Reliability. *Behaviormetrika*, Vol.40, No.2, 149–168

このとき、

$$P(X_i = 1; \theta) = P(Z_i > 0; \theta) = \Phi\left(\frac{a_i + b_i\theta}{\sigma}\right)$$

$$P(X; \theta) = \sum_{X=\sum X_i} P(X_i; \theta)$$

である。これより、

$$P(\theta, X) = P(\theta)P(X; \theta)$$

を得る。

$P(X; \theta)$ はシミュレーションで推定できる。 $P(\theta)$ は、

$$P(\theta) = N(\theta; 0, 1)$$

の場合について考える。

次式の積分

$$P(X) = \int_{\theta} P(\theta, X) d\theta$$

は、 θ を離散化して求めた。これより

$$P(\theta|X) = \frac{P(\theta, X)}{P(X)}$$

を得る。

不確かさを分散で与える。

$$U(\theta) = \text{Var}(\theta),$$

$$U(X) = \text{Var}(X),$$

$$U(\theta|X) = \text{Var}(\theta|X),$$

$$U(X|\theta) = \text{Var}(X|\theta)$$

このとき、信頼性は次式で与えられる。

$$R_c(\theta) = \frac{\text{Var}(X) - \text{Var}(X|\theta)}{\text{Var}(X)}$$

$$R_m(X) = \frac{\text{Var}(\theta) - \text{Var}(\theta|X)}{\text{Var}(\theta)}$$

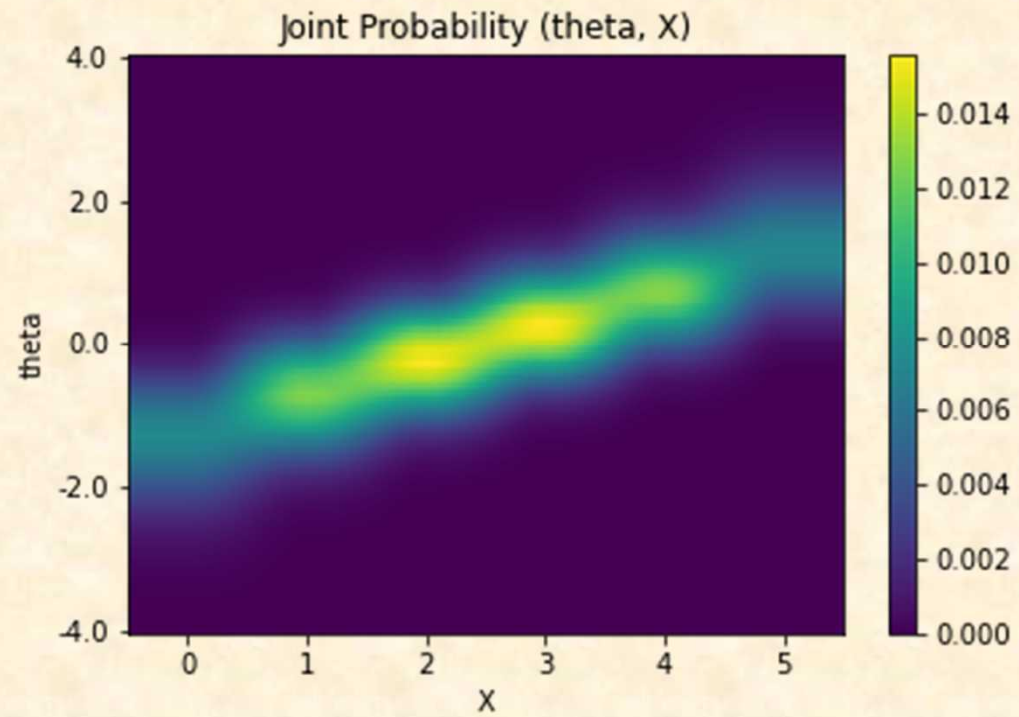
例として、次の順序カテゴリ尺度を考える。

項目数 $M = 5$

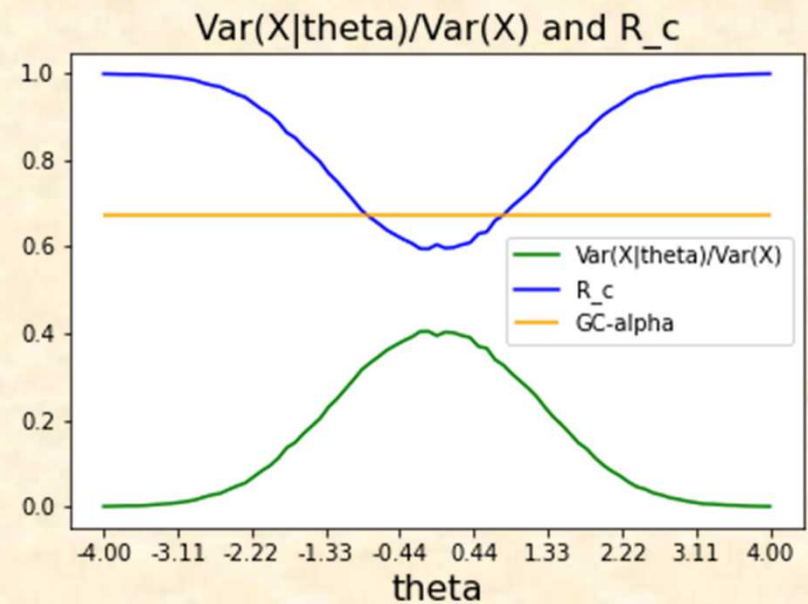
係数 $a_i = -1, -0.5, 0, +0.5, +1$ 、 $b_i = 1$

$\sigma = 1$

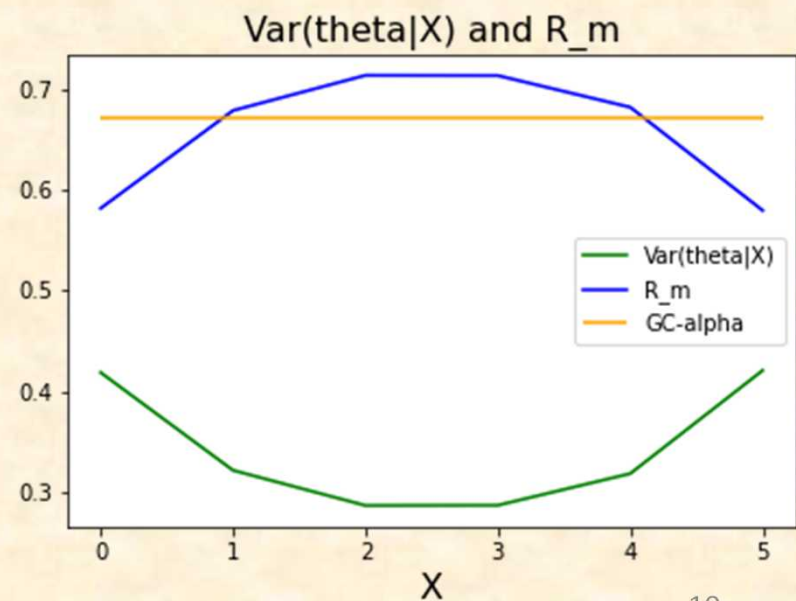
同時確率 $P(\theta, X) = P(\theta)P(X; \theta)$



$$\text{信頼性係数 } R_c(\theta) = \frac{\text{Var}(X) - \text{Var}(X|\theta)}{\text{Var}(X)}$$



$$\text{信頼性係数 } R_m(X) = \frac{\text{Var}(\theta) - \text{Var}(\theta|X)}{\text{Var}(\theta)}$$



項目反応理論

概念値 θ に対して各項目 X_i の確率分布が与えられ、回答パターン $\{X_i\}$ （CATでは項目群は回答に応じて決められる）に対して θ の推定値 $\hat{\theta}$ が測定値として与えられる。測定値を X ではなく $\hat{\theta}$ で表す。

順序カテゴリ尺度のモデルを次式の形に書き換える。

$$P(X_i = 1; \theta) = \Phi\left(\frac{a_i + b_i\theta}{\sigma}\right) = \Phi(\alpha_i(\theta - \beta_i)) \quad (\text{IRT.0})$$

ここで、

$$\alpha_i = \frac{b_i}{\sigma}, \quad \beta_i = -\frac{a_i}{b_i}$$

である。モデル（IRT.0）は、2パラメータ正規モデルと呼ばれている（岡本、2014）。

岡本安晴（2014）．心理学データ分析と測定．培風館

式 (IRT.0) より、測定対象概念値 θ に対して各項目 X_i の確率が与えられ項目の回答パターン $\{X_i\}$ に対する確率が以下のように与えられる。

$$P(\{X_i\}|\theta) = \prod_{i=1}^M P(X_i; \theta) \quad (\text{IRT. 1})$$

回答パターン $\{X_i\}$ と測定対象概念値 θ の事前分布 $P(\theta)$ から事後分布が次式で与えられる。

$$P(\theta|\{X_i\}) \propto P(\theta)P(\{X_i\}|\theta) \quad (\text{IRT. 2})$$

ここで、

$$P(\theta) = N(\theta; 0,1) \quad (\text{IRT. 3})$$

の場合について考える。

事後分布 (IRT.2) から θ の推定値を求めるが、ここでは事後分布の中央値を推定値とする。

$$\hat{\theta} = \text{Arg med}_{\theta}(P(\theta|\{X_i\}))$$

以上より

$$P(\hat{\theta}|\theta) = \sum_{\hat{\theta}=\text{med}(P(\theta|\{X_i\}))} P(\{X_i\}|\theta) \quad (\text{IRT. 4})$$

を得る。 $P(\hat{\theta}|\theta)$ はシミュレーションで推定することが出来る。なお、モデル

(IRT.1) では項目群 $\{X_i\}$ が固定されているが、適応的方法のようにテスト受験者の回答に応じて項目 X_i が選ばれる場合も、シミュレーションによって式 (IRT.4) の推定を行うことが出来る。

式 (IRT.3) と (IRT.4) より同時確率

$$P(\theta, \hat{\theta}) = P(\theta)P(\hat{\theta}|\theta) \quad (\text{IRT.5})$$

を得ることが出来る。式 (IRT.5) より

$$P(\hat{\theta}) = \int_{-\infty}^{+\infty} P(\theta, \hat{\theta})d\theta, \quad P(\theta|\hat{\theta}) = \frac{P(\theta, \hat{\theta})}{P(\hat{\theta})}$$

を得る。

いま、不確かさを分散で与える。

$$U(\theta) = \text{Var}(\theta), \quad U(\theta|\hat{\theta}) = \text{Var}(\theta|\hat{\theta}),$$

$$U(\hat{\theta}) = \text{Var}(\hat{\theta}), \quad U(\hat{\theta}|\theta) = \text{Var}(\hat{\theta}|\theta)$$

したがって、信頼性は次式で与えられる。

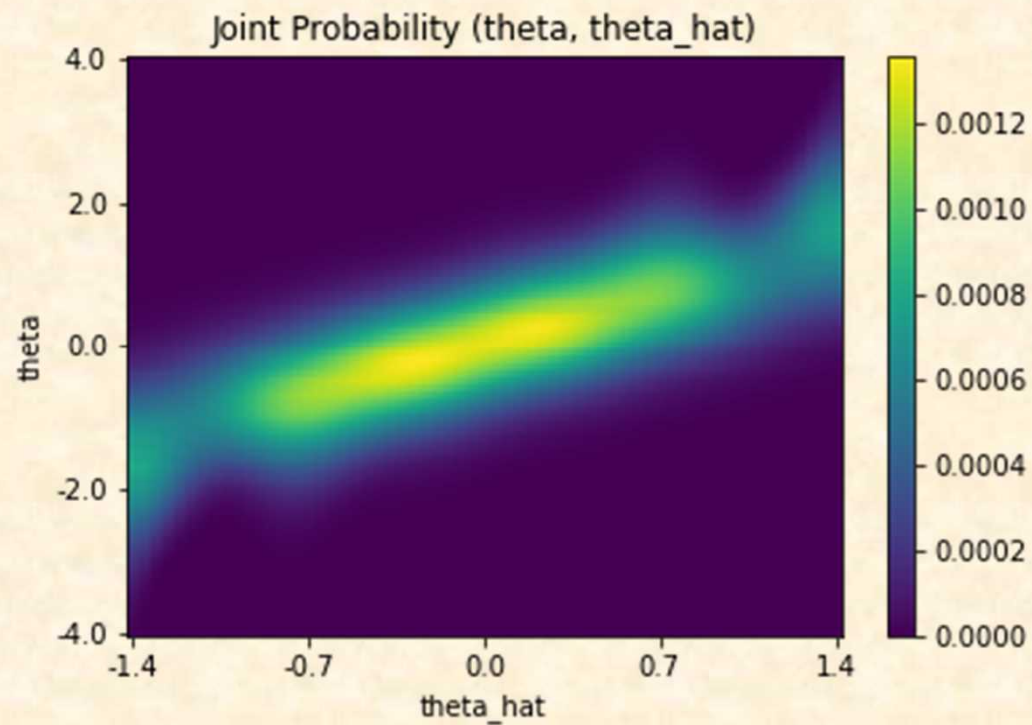
$$R_c(\theta) = \frac{U(\hat{\theta}) - U(\hat{\theta}|\theta)}{U(\hat{\theta})} = \frac{\text{Var}(\hat{\theta}) - \text{Var}(\hat{\theta}|\theta)}{\text{Var}(\hat{\theta})}$$

$$R_m(\hat{\theta}) = \frac{U(\theta) - U(\theta|\hat{\theta})}{U(\theta)} = \frac{\text{Var}(\theta) - \text{Var}(\theta|\hat{\theta})}{\text{Var}(\theta)}$$

項目反応理論の場合の信頼性係数を、上の順序カテゴリ尺度の例におけるパラメータ値に対して求めてみる。すなわち、

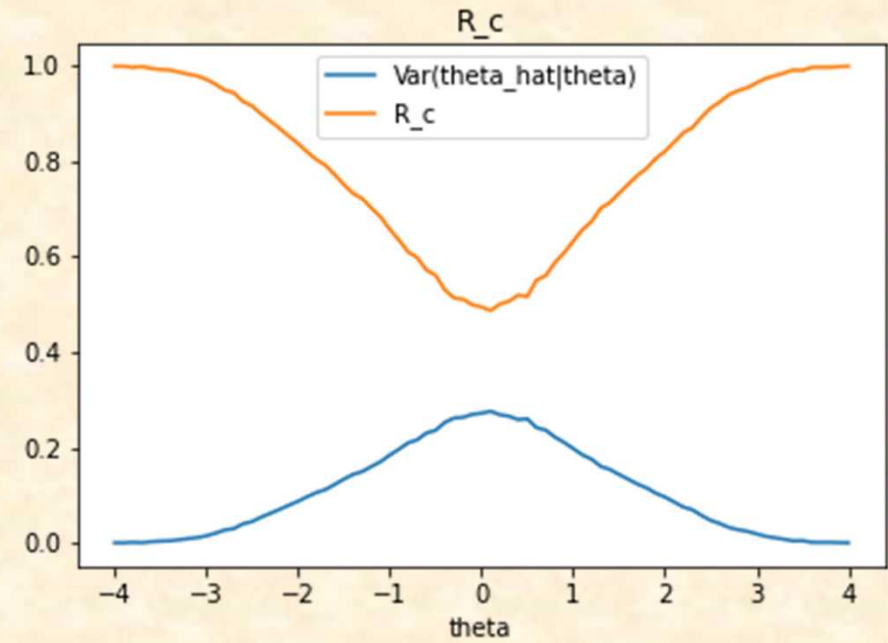
項目数 $M = 5$ 、係数 $a_i = -1, -0.5, 0, +0.5, +1$ 、 $b_i = 1$ 、 $\sigma = 1$ と置く。

同時確率 $P(\theta, \hat{\theta})$



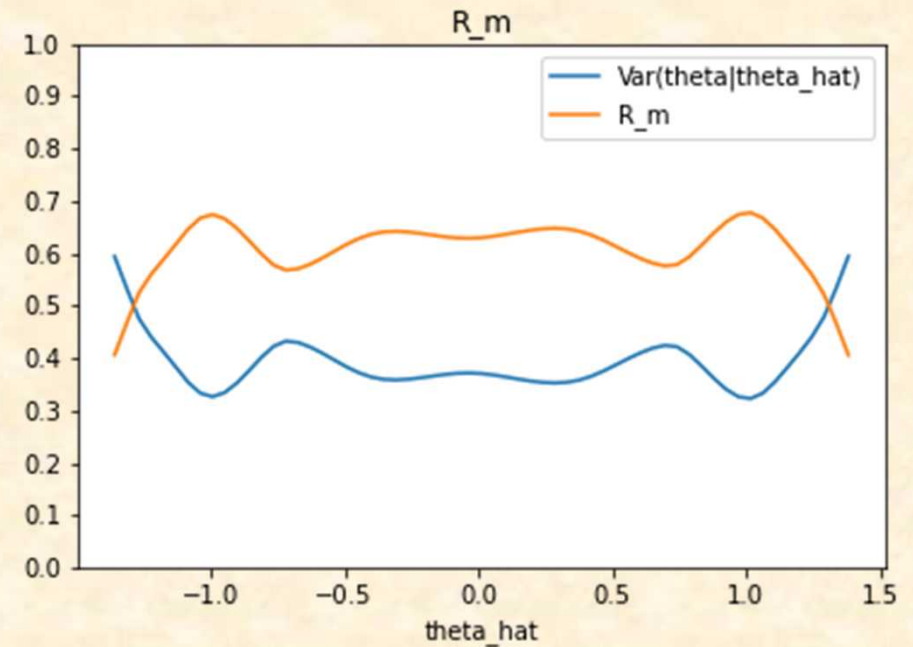
信頼性係数 $R_c(\theta)$

$$= \frac{U(\hat{\theta}) - U(\hat{\theta}|\theta)}{U(\hat{\theta})} = \frac{Var(\hat{\theta}) - Var(\hat{\theta}|\theta)}{Var(\hat{\theta})}$$



信頼性係数 $R_m(\hat{\theta})$

$$= \frac{U(\theta) - U(\theta|\hat{\theta})}{U(\theta)} = \frac{Var(\theta) - Var(\theta|\hat{\theta})}{Var(\theta)}$$



潜在クラス尺度

潜在クラス C が概念値であり、推定潜在クラス \hat{C} が測定値である。

潜在クラスが K 個あるとする。潜在クラス C における M 個の項目 X_i に対する回答確率が $P(X_i|C)$ であるとき、潜在クラス C における M 個の回答パターン $X = \{X_i\}$ を次式で与える。

$$P(X|C) = \prod_i P(X_i|C)$$

このとき、同時確率は次式で与えられる。

$$P(C, X) = P(C)P(X|C)$$

これより、次式を得る。

$$P(C|X) = \frac{P(C, X)}{P(X)} = \frac{P(C)P(X|C)}{P(X)} \quad (\text{LC. 1})$$

反応パターン $\{X_i\}$ に対する推定潜在クラス \hat{C} を事後分布 (LC.1) のモードとして求めることにする (MAP推定値)。

$$\hat{C} = \underset{C}{\text{Arg max}} P(C|X)$$

確率 $P(\hat{C}|C)$ はシミュレーションで推定できる。これより、同時確率は次式で与えられる。

$$P(C, \hat{C}) = P(C)P(\hat{C}|C)$$

これより次式を得る。

$$P(C|\hat{C}) = \frac{P(C, \hat{C})}{P(\hat{C})}, \quad P(\hat{C}) = \sum_C P(C, \hat{C})$$

クラスの確率分布に対する不確かさをエントロピーで与える。ここで、確率分布 $P = \{p_i\}$ に対してエントロピーは次式で与えられる (Applebaum, 1996)。

$$Entropy(P) = - \sum p_i \log_2 p_i$$

このエントロピーにより

$$\begin{aligned} U(C) &= Entropy(P(C)) \\ U(C|\hat{C}) &= Entropy(P(C|\hat{C})) \\ U(\hat{C}) &= Entropy(P(\hat{C})) \\ U(\hat{C}|C) &= Entropy(P(\hat{C}|C)) \end{aligned}$$

とおく。

Applebaum, D. (1996). *Probability and information: An integrated approach*. Cambridge: Cambridge University Press.

このとき、信頼性係数は次式で与えられる。

$$R_c(C) = \frac{U(\hat{C}) - U(\hat{C}|C)}{U(\hat{C})}, \quad R_m(\hat{C}) = \frac{U(C) - U(C|\hat{C})}{U(C)}$$

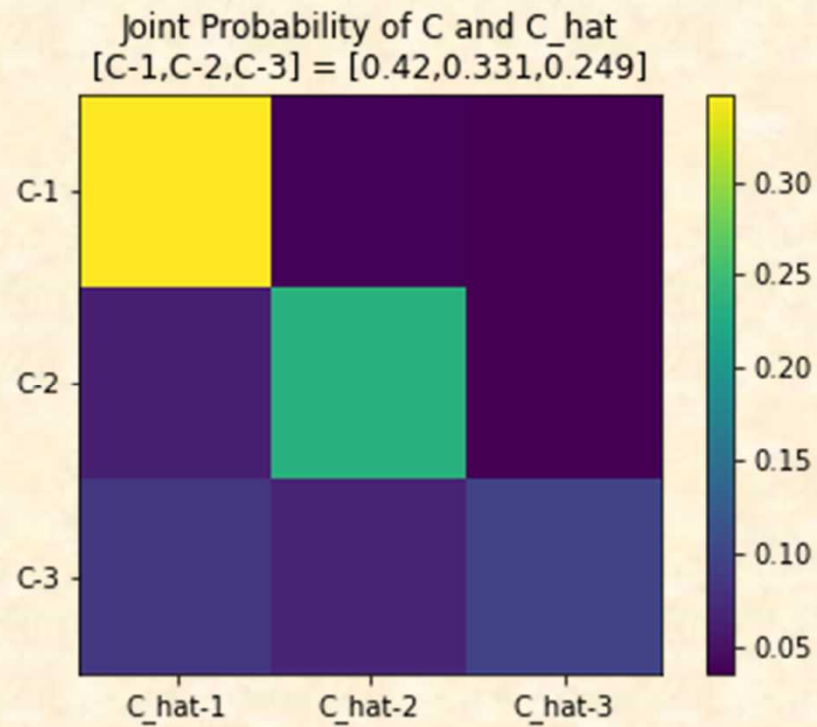
潜在クラス尺度の信頼性係数を、稲垣・前田（2015）の尺度に対して求めてみる。

稲垣・前田（2015）の表3では3つの潜在クラス、クラス1（用心深さ）、クラス2（一般的信頼）、クラス3（不信（安心））における、5つの質問項目、「他人か自分か」、「利用されるか」、「人は信頼できるか」、「しきたりに従うか」、「スジかまるくか」に対する応答確率が報告されている。質問項目に対する回答は、「しきたりに従うか」が3選択肢、他の項目は2選択肢であった。表3における第1選択肢への回答確率を共変量投入後の推定値をTable 1にまとめた。

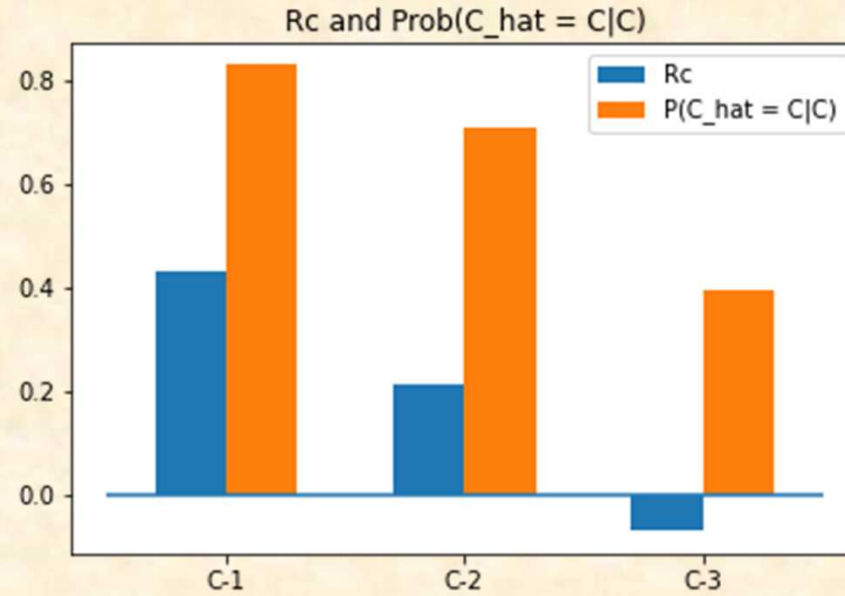
稲垣祐典・前田忠彦（2015）潜在クラス分析による「日本人の国民性調査」における信頼の意味とその時代的変遷の検討. 統計数理、63、277-297.

Table 1						
		質問項目				
潜在クラス	構成割合	Q1	Q2	Q3	Q4	Q5
C-1	0.420	0.859	0.607	0.859	0.357	0.430
C-2	0.331	0.464	0.094	0.348	0.274	0.470
C-3	0.249	0.598	0.198	0.717	0.632	0.707

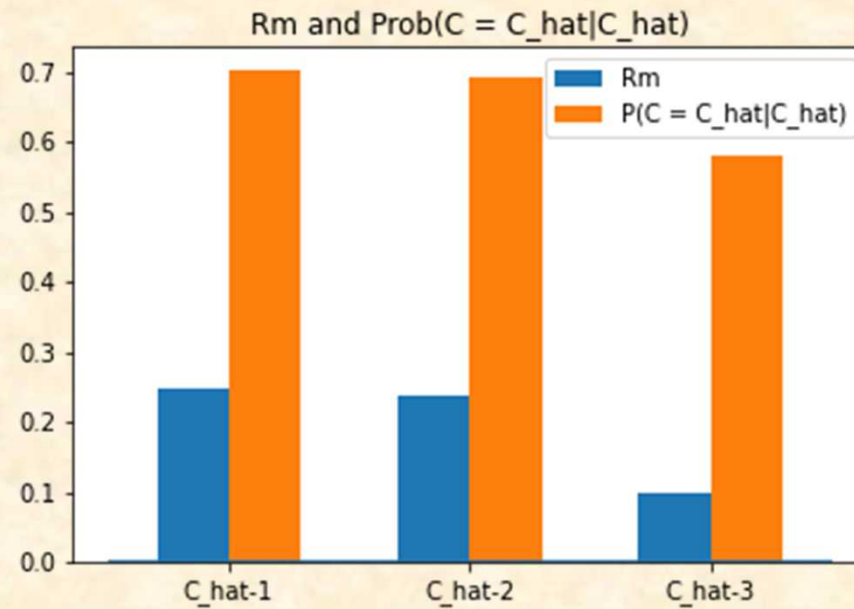
Table 1の場合の同時確率分布 $P(C, \hat{C})$ は下図のようである。



$$\begin{aligned} \text{信頼性係数 } R_c(C) \\ &= \frac{U(\hat{C}) - U(\hat{C}|C)}{U(\hat{C})} \end{aligned}$$

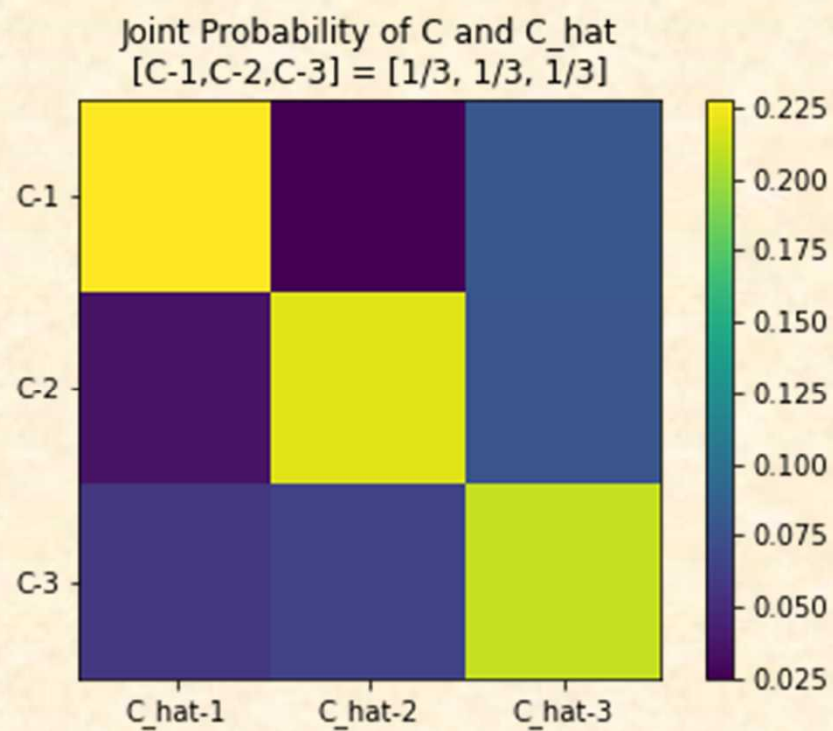


$$\begin{aligned} \text{信頼性係数 } R_m(\hat{C}) \\ &= \frac{U(C) - U(C|\hat{C})}{U(C)} \end{aligned}$$

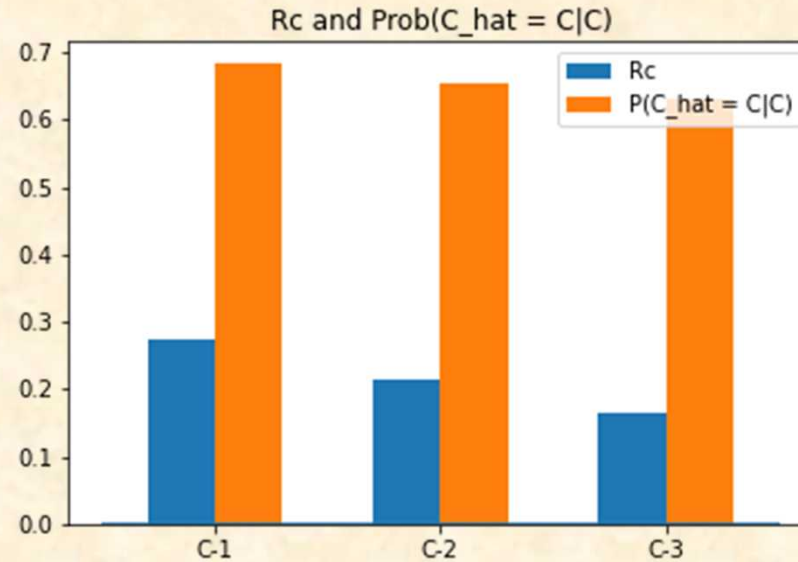


事前確率分布 $P(C = 1) = P(C = 2) = P(C = 3) = 1/3$ の場合

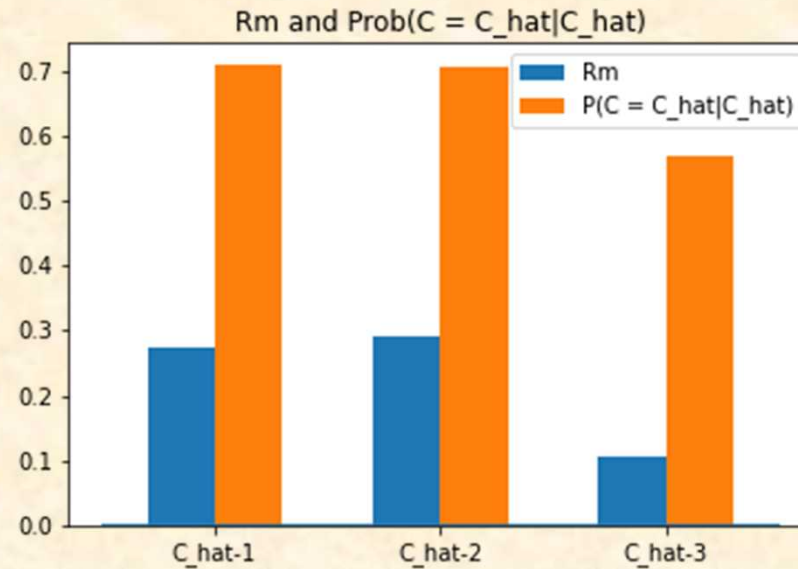
同時確率分布 $P(C, \hat{C})$



$$\begin{aligned} \text{信頼性係数 } R_c(C) & \\ &= \frac{U(\hat{C}) - U(\hat{C}|C)}{U(\hat{C})} \end{aligned}$$



$$\begin{aligned} \text{信頼性係数 } R_m(\hat{C}) & \\ &= \frac{U(C) - U(C|\hat{C})}{U(C)} \end{aligned}$$



まとめ

1. 誤差モデルに基づくのではなく、不確かさに基づいて信頼性を考えることを提案した。不確かさに基づくと、値がクラスという離散的な場合も含めて統一的に考えることができる。また、概念値から見た不確かさの係数 R_c と、測定値から見た不確かさの係数 T_m の2方向が区別された。
2. 古典テスト理論の場合、不確かさに基づく信頼性係数は両方向とも誤差理論に基づく信頼性係数と同じ値であることが示された。
3. 順序カテゴリ尺度、最も広く用いられている尺度と思われるが、簡単な場合についてシミュレーションにより検討した。 R_c は中ほどで低く、 R_m は中ほどが高いことが示された。 α 係数は、これらの値の中ほどであった。
4. 項目反応理論を上記の順序カテゴリ尺度の項目パラメータと同じものを設定して検討を行った。 R_c については順序カテゴリ尺度の場合と同様であったが、 R_m については中ほどでの値の増加は見られず、中ほどでは順序カテゴリ尺度より少し低い値であった。
5. 潜在クラス尺度は、学会誌に報告されているものを用いて計算した。 R_c が負の値のものが出たが、事前分布を一様分布とすると正の値になった。事前分布を踏まえた信頼性の検討の重要性を示すものと言える。