

予測信頼性係数

○岡本安晴

日本女子大学人間社会学部

1. はじめに

テスト得点の精度は信頼性と呼ばれ、テスト得点の分散に占める真値の分散の比率として信頼性係数が定義される (McDonald, 1999)。しかし、この信頼性係数は、真値 T とテスト得点 X の関係が

$$X = T + E, \quad E: \text{誤差} \quad (1)$$

という形に表されたときに定義されるものである。回答がカテゴリー値 Y の場合は、(1) 式の形ではなく、例えば ordinal probit regression が用いられる (Okamoto, 2013)。信頼性を、回答得点の和 Y と真値 T との関係と考えたとき、次の回帰モデル

$$T = aY + b + e, \quad e: \text{残差} \quad (2)$$

に基づく決定係数によって表すことが考えられ、岡本 (2012) は予測的信頼性係数として提案した。これを本報告では予測信頼性係数 (predictive reliability coefficient) ρ_p^2 と呼び、ベイズ的分析法によって求めたときの特性を調べる。なお、テスト得点が連続量の場合 (式 (1))、決定係数は信頼性係数に一致するので、決定係数をベイズ的分析で求めたときの特性は、信頼性係数をベイズ的分析によって求めたときの特性であり、これは Okamoto (2013) などに報告されている。本報告では、回答がカテゴリー変数の場合について、ベイズ的分析法で求めた予測信頼性係数の特性をシミュレーションによって調べた結果を報告する。

2. 予測信頼性係数

測定値 Y によって測定対象の値 T を予測する式 (2) を設定したとき、 Y による予測値 $aY + b$ の分散の測定対象の値 T の分散に対する比は決定係数と呼ばれる。これは Y の式 (2) による T の予測の精度を表すものであり、 Y の信頼性の指標となる。これを予測信頼性係数 ρ_p^2 と呼べば、決定係数は相関係数の 2 乗で与えられることから

$$\rho_p^2 = (\rho_{YT})^2, \quad (3)$$

が成り立つ。ここで、

$$\rho_{YT} = Y \text{ と } T \text{ の相関係数}$$

である。

式 (3) において、 Y は測定値であるが T は潜在変数であるので、 ρ_{YT} を求めるときは T の推定値 $\hat{T}$ に基づくことが考えられる。このときの ρ_{YT} の値を $r_{Y\hat{T}}$ 、 ρ_p^2 の値を $\hat{\rho}_p^2$ とおけば

$$\hat{\rho}_p^2 = (r_{Y\hat{T}})^2$$

となる。

尺度項目が m 個のカテゴリー変数からなるとき、回答者 i の尺度値が

$$Y_i = \sum_{j=1}^m Y_{ij}$$

で与えられるとする。ここで、 Y_{ij} は回答者 i のカテゴリ変数 j に対するカテゴリ回答である。

回答者 i の真値 T_i の推定値 $\hat{T}_i$ は、ベイズ的方法によって求める場合は回答カテゴリのパターン $Y_{i1}, \dots, Y_{im}$ に基づく事後分布 $P(T_i | Y_{i1}, \dots, Y_{im})$ で与えられる。この事後分布から $\hat{\rho}_p^2$ の事後分布 $P(\hat{\rho}_p^2 | Y_{i1}, \dots, Y_{im})$ を求めることができる。Okamoto (2013)では、 $Y_{i1}, \dots, Y_{im}$ の回答パターンに対するベイズ的分析は、回答カテゴリ数が2の場合と3以上の場合とでモデル設定における制約条件は異なったものが用いられている。これは、回答カテゴリ数が3以上の場合はカテゴリ境界の数は2以上であるが、回答カテゴリ数が2の場合はカテゴリ境界の数は1となり、回答カテゴリ数が3以上の場合と2の場合とでカテゴリ境界による潜在変数 X の原点と単位に対する制約の掛かり方が異なるためである。本報告でも基本モデルはOkamoto (2013)に従い、回答カテゴリ数が2の場合と3以上の場合に分けて扱う。

3. 回答カテゴリ数が3以上の場合

項目 j に対する回答者 i のカテゴリ回答を Y_{ij} 、その潜在変数を X_{ij} とする。カテゴリ境界を C_k で表したとき、 Y_{ij} がカテゴリ k であるのは $C_{k-1} \leq X_{ij} < C_k$ のときであるとする。すなわち、

$$C_{k-1} \leq X_{ij} < C_k \quad \text{のとき} \quad Y_{ij} = k \quad (4)$$

ここで、カテゴリ数を K とすると

$$-\infty = C_0 < C_1 < \dots < C_{K-1} < C_K = +\infty$$

であるとする。

潜在変数 X_{ij} は、回答者 i の特性 F_i によって次式

$$X_{ij} = \mu_j + \lambda_j F_i + E_{ij}, \quad E_{ij} \sim N(0, \sigma_{Ej}^2) \quad (5)$$

で与えられるとする。ここで、 $F_i \sim N(0, 1)$ とする。 $T_{ij} = \mu_j + \lambda_j F_i$ が真値、 E_{ij} が誤差であるが、 F_i が回答者の個人差(特性)を表す真値であると考え、回答者の尺度値

$$Y_i = \sum_j Y_{ij} \quad (6)$$

と真値 T すなわち F_i との相関係数を求め、その2乗の値として決定係数を算出して予測信頼性係数とする。この予測信頼性係数のベイズ的分析の特性を以下のシミュレーションによって調べる。

- (a) 回答者 i の真値 $F_i \sim N(0, 1); i = 1, \dots, n$ を生成し、式(5)に従って潜在変数 $X_{ij}; j = 1, \dots, m$ の値を設定する。 X_{ij} の値から式(4)によってカテゴリ回答 Y_{ij} を設定し、尺度値 Y_i を式(6)によって求める。 F_i と Y_i の相関係数の2乗を真の予測信頼性係数 ρ_p^2 とする。

シミュレーションによって生成されたカテゴリ回答データ Y_{ij} に対して、事後分布を次のMCMCにより求める。

- (b) ステップ t におけるパラメータのサンプル値を $F_i^{(t+1)}, \mu_j^{(t+1)}, \lambda_j^{(t+1)}, C_k^{(t+1)}, \sigma_{Ej}^{(t+1)}$ の順にOkamoto (2013)で用いられているMetropolis-within-Gibbsによって求める。 $F_i^{(t+1)}$

と Y_i との相関係数の 2 乗を求め、予測信頼性係数のステップ t におけるサンプル値 $R2^{(t+1)}$ とする。

ステップ (b) における MCMC によるサンプリングを $Nmcmc$ 回繰り返す。

(c) $Nmcmc$ 個の $R2^{(t)}$ のサンプル値 ($t = 1, \dots, Nmcmc$) を MCMC による事後分布のサンプルとする。このサンプルにより事後分布の credible interval (CI) などを調べる。

回答選択肢数が 7 カテゴリーの項目を 10 項目用意し、パラメータ値を

$$\mu_1 = \dots = \mu_{10} = 0, \quad \lambda_1 = \dots = \lambda_5 = 0.8, \quad \lambda_6 = \dots = \lambda_{10} = 0.4$$

$$\sigma_{E1} = \dots = \sigma_{E5} = 0.6, \quad \sigma_{E6} = \dots = \sigma_{E10} = 0.92$$

$$C_1 = -1.23, C_2 = -0.738, C_3 = -0.246, C_4 = 0.246, C_5 = 0.738, C_6 = 1.23$$

とおき、200 人の回答者を想定して (a) ~ (c) のシミュレーションを $Nsimu = 300$ 回繰り返し、各シミュレーションにおいて (c) で求めた予測信頼性係数の 95% CI が (a) で算出された真値 ρ_p^2 を含む回数を数えたところ 284 回であった。284/300 $\approx$ 0.947 であるので、上の MCMC による 95% CI の推定は満足のいくものであるといえる。

4. 回答カテゴリ数が 2 の場合

カテゴリ数 $K = 2$ のときは、カテゴリ境界が 1 個であり、Okamoto (2013) は信頼性係数の算出においてモデルに essentially tau-equivalent の制約を置き、さらに分散の推定が不安定になることを避けるために parallel の条件を加えている。本報告においても essentially tau-equivalent の場合を扱うが、Okamoto (2013) における解の安定性を高めるため parallel の条件に加えて分散の事前分布として inverse-gamma 分布を用いる。Inverse-gamma 分布は 2 つのパラメータ α と β によって $\text{Inv-gamma}(\alpha, \beta)$ と表されるが、 $\alpha = 2.01$ 、 $\beta = 3$ のものを用いる。 $\text{Inv-gamma}(2.01, 3)$ のモードは $\beta/(\alpha + 1) \approx 0.997$ 、標準偏差は $\sqrt{\beta^2/\{(\alpha - 1)^2(\alpha - 2)\}} \approx 29.703$ である (図 1)。分散の事前分布として inverse-gamma 分布を用いること以外は Okamoto (2013) における Metropolis-within-Gibbs による方法と同じものを用いる。すなわち、

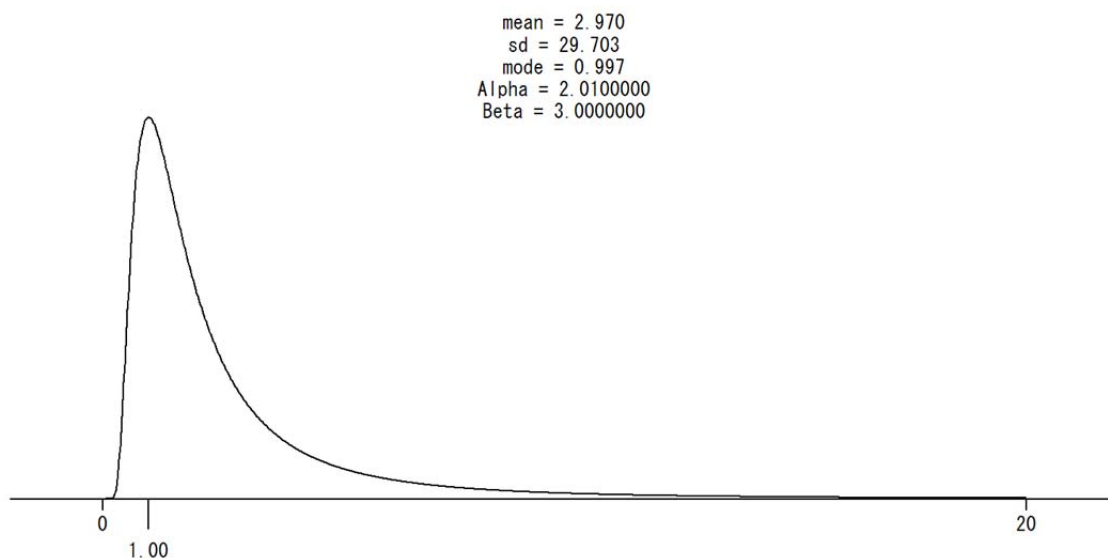


図 1 Inverse-gamma 分布

以下の手順によりシミュレーションを行う。

- (a) 回答者 i の真値 $F_i \sim N(0, 1); i = 1, \dots, n$ を生成し、式 (5) に従って潜在変数 $X_{ij}; j = 1, \dots, m$ の値を設定する。 X_{ij} の値から式 (4) によってカテゴリ回答 Y_{ij} を設定し、尺度値 Y_i を式 (6) によって求める。 F_i と Y_i の相関係数の 2 乗を真の予測信頼性係数 ρ_p^2 とする。

カテゴリ回答データ Y_{ij} に対して事後分布を MCMC により求める。

- (b) ステップ t におけるパラメータのサンプル値を $F_i^{(t+1)}$ 、 $\mu_j^{(t+1)}$ 、 $\sigma_E^{(t+1)}$ の順に Okamoto (2013) で用いられている Metropolis-within-Gibbs によって行う。Essentially tau-equivalent 項目を仮定しているので $\lambda_j = 1$ とおき、parallel の条件より $\sigma_{Ej}^{(t+1)}$ は項目 j によらず同じ値 $\sigma_E^{(t+1)}$ とする。 $F_i^{(t+1)}$ と Y_i との相関係数の 2 乗を求め、予測信頼性係数のステップ t におけるサンプル値 $R2^{(t+1)}$ とする。

ステップ (b) における MCMC によるサンプリングを Nmcmc 回繰り返す。

- (c) Nmcmc 個の $R2^{(t)}$ のサンプル値 ($t = 1, \dots, Nmcmc$) を MCMC による事後分布のサンプルとする。このサンプルにより事後分布の credible interval (CI) などを調べる。

回答選択肢数が 2 カテゴリーの項目を 10 項目用意し、パラメータ値を

$$\begin{aligned}\mu_1 = \dots = \mu_{10} = 0, \quad \lambda_1 = \dots = \lambda_{10} = 0.7 \\ \sigma_{E1} = \dots = \sigma_{E10} = \sqrt{1 - 0.7^2}, \quad C_1 = 0\end{aligned}$$

とおき、200 人の回答者を想定して (a) ~ (c) のシミュレーションを Nsimu = 300 回繰り返し、各シミュレーションにおいて (c) で求めた予測信頼性係数の 95% CI が (a) で算出された真値 ρ_p^2 を含む回数を数えたところ 290 回であった。290/300 = 0.967 であるので、上の MCMC による 95% CI の推定は満足のいくものであるといえる。

5. まとめ

測定値は真値との関係が強いほど信頼性があると考えられる。測定値が真値と同じ次元上の数値として考えられる連続変量の測定値の場合は、測定値に占める真値の割合すなわち精度として信頼性を表すことができるが、カテゴリ値によって測定される場合は、測定値と真値は異なる次元に属するものとなり、測定値に占める真値の割合という考え方は不適切である。カテゴリ値の場合は、項目和としての尺度値と真値との関係を表すものとして信頼性を考え、決定係数を信頼性係数として用いることを提案した。この予測信頼性係数をベイズ的に推定する場合のシミュレーションを、7 カテゴリーの場合と 2 カテゴリーの場合について行った。それぞれ 1 つの条件でのシミュレーションであるが、95%CI は約 95% の確率で予測信頼性係数を含んでおり、その有用性が示唆された。

引用文献

- 岡本安晴 (2012). 日本テスト学会第 10 回発表論文抄録集、164-167.
Okamoto, Y. (2013). *Behaviormetrika*, 149-168.