

予測アセスメントと相関係数

関連特性の予測確率

岡本安晴

日本女子大学

相関係数は、因子分析あるいは共分散構造分析において基礎となる重要な統計量である。

しかし、2つの心理特性の関係の日常的推論においては、相関係数に基づく単純な評価（悉無率的予測）は危険である。具体的な数値例を以下に報告する。

まず、個人内の推論（予測）について示し、次に個人間の推論（予測）を扱う。

個人内の推論（予測）

特性**A**と**B**があり、**A**と**B**の間には相関があるとする。
相関係数 ρ の値を正、各特性の平均を0、分散を1と
仮定しても、以下の議論の一般性は失われない。
すなわち、

$$(1) \quad \begin{bmatrix} A \\ B \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

とする。このとき、

「特性**A**のある人は特性**B**がある」

と悉無率的に言うことは不適切である。

相関係数は確率的関係を表すものであり、
「特性Aのある人は特性Bがある」
という決定論的言明ではない。

相関係数に基づいた推論では、この言明の
成り立つ確率を考えなければならない。

言明「特性Aのある人は特性Bがある」
の確率は

$$P(B > 0 | A > 0)$$

で表される。この確率を相関係数 ρ の関数として
描いたものを図1に示す。

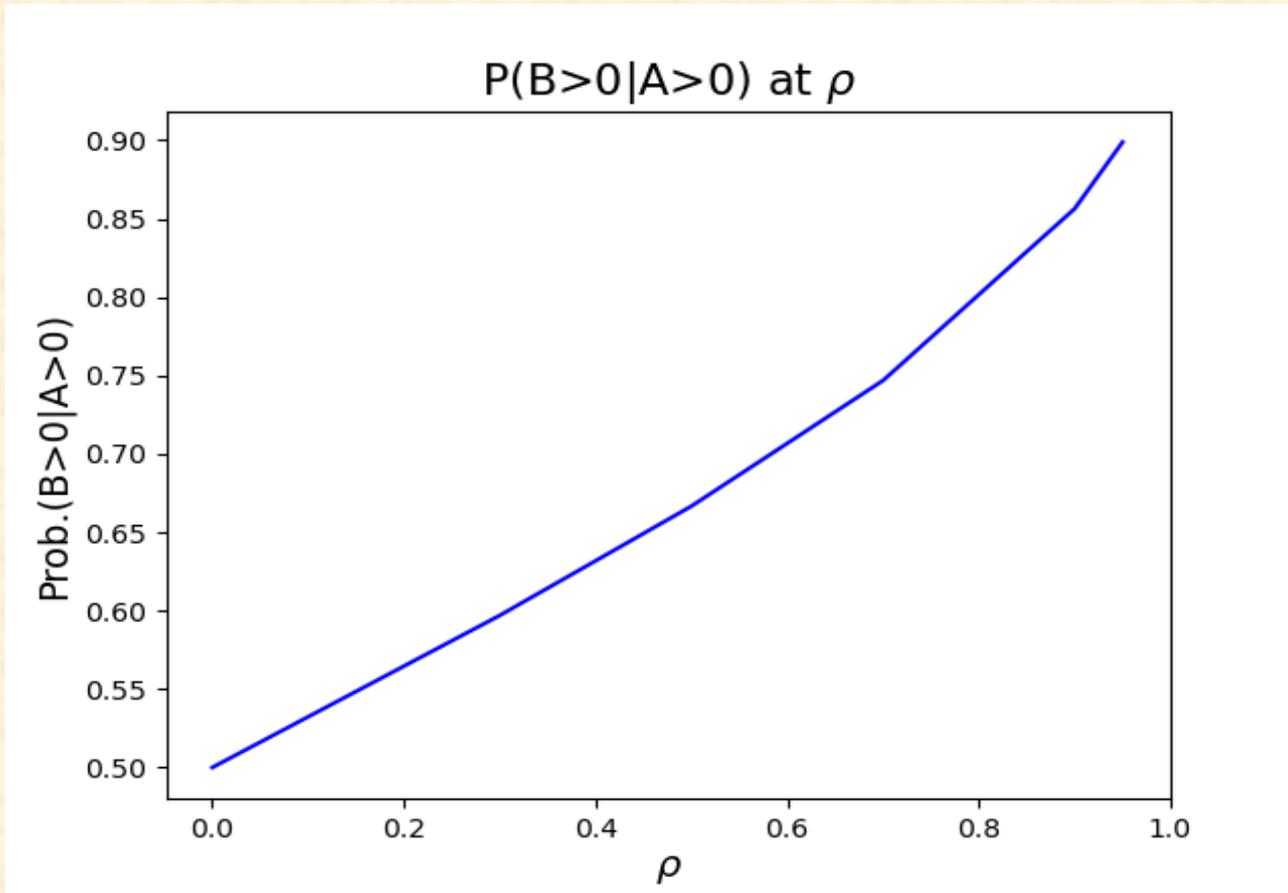


図1 確率 $\text{Prob}(B > 0 | A > 0)$ と相関係数 ρ

相関が無い ($\rho = 0$) とき、
言明「特性Aのある人は特性Bがある」の成り立つ
確率は0.5で、言明の当たりはずれの確率は等しい。

相関係数が $\rho = 0.5$ のときは、
言明「特性Aのある人は特性Bがある」の成り立つ
確率は約2/3で、3人に1人の割合で外れる。

相関係数が0.5以下のときは、
言明「特性Aのある人は特性Bがある」のはずれは
無視できない確率である。

相関係数が0.8を超えると
言明の外れる確率も5人に1人以下になる。

言明（予測）の成り立つ確率は、
Aの値が大きい正の値であるほど
高くなると考えられる。

Aの値の関数として確率

$$P(B > 0|A)$$

を描くと図2のようになる。

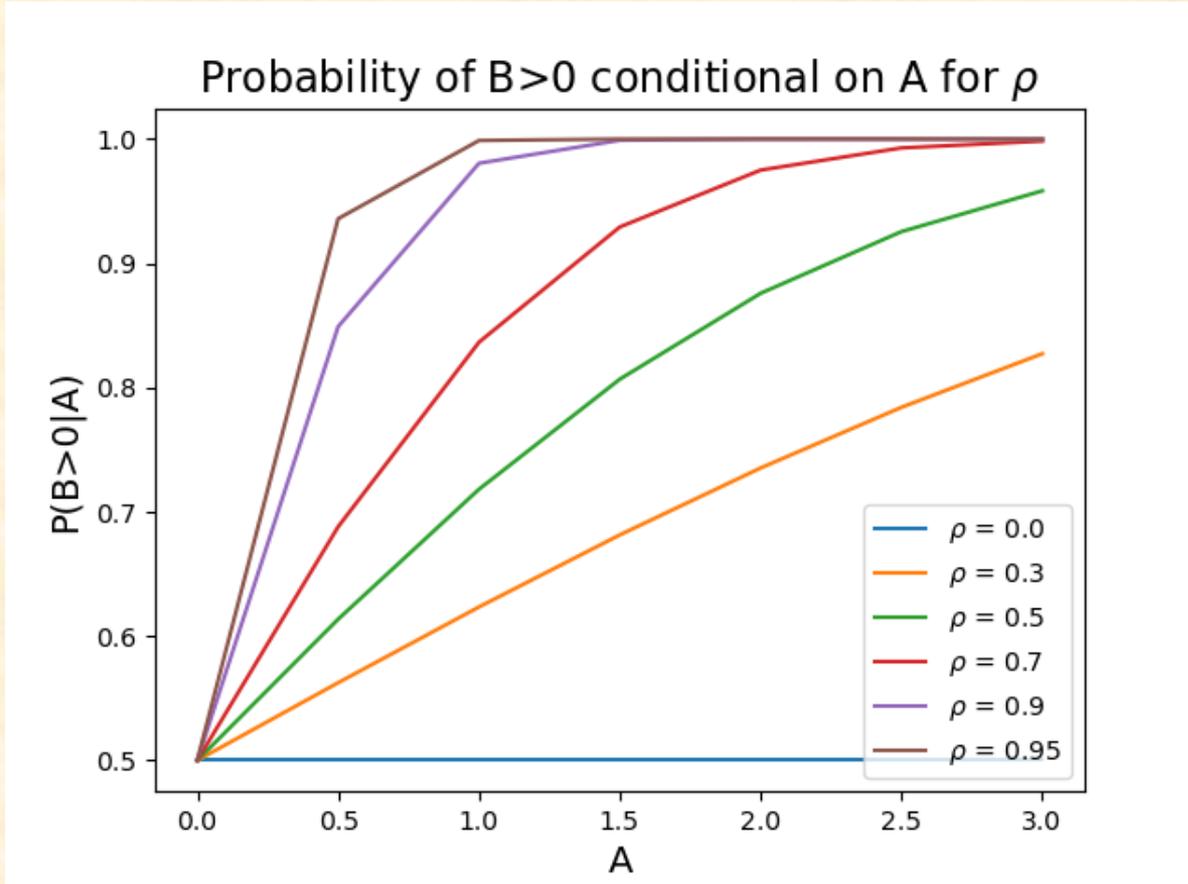


図2 確率 $P(B > 0 | A)$ と A の値

相関係数 ρ が0の場合は、
Aの値に関わらず言明「特性Aのある人は特性Bがある」が
成り立つ確率は0.5である。

相関係数 ρ が0.9のとき、
Aが1.5以上であれば言明が成り立つ確率は1.0である
と言える。

相関係数が0.3であれば、
Aが2.0のときで4人に1人ぐらいの割合で
言明「特性Aのある人は特性Bがある」は成り立たない。

個人間の推論

個人 p と個人 q の比較の場合について調べる。

それぞれの値を $[A_p, B_p]^T$ 、 $[A_q, B_q]^T$ で表すと、

独立性の仮定の下で、

$$(2) \quad \begin{pmatrix} A_p - A_q \\ B_p - B_q \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 2\rho \\ 2\rho & 2 \end{bmatrix} \right)$$

となる

個人 p の方が個人 q より特性 A が強い場合、
特性 B においても個人 p の方が特性が強い確率

$$P(B_p - B_q > 0 | A_p - A_q > 0)$$

は、相関係数 ρ の関数として図1のグラフで表される。

これは、式(1)と式(2)を比較して
式(2)は式(1)の確率分布を原点を中心として
 $\sqrt{2}$ 倍に引き延ばしたものであり、求める確率は
領域

$$\{B_p - B_q > 0, A_p - A_q > 0\}$$

と領域

$$\{A_p - A_q > 0\}$$

の確率の比であることから分かる。

相関が無い ($\rho = 0$) 場合、
「個人 p の方が特性Aが強い」とき、
予測「個人 p の方が特性Bが強い」
の成り立つ確率は0.5で、予測の当たりはずれの確率は等しい。

相関係数が $\rho = 0.5$ のときは、
予測「個人 p の方が特性Bが強い」の成り立つ確率は約2/3で、
3人に1人の割合で外れる。

相関係数が0.5以下のときは、
予測「個人 p の方が特性Bが強い」
のはずれは無視できない確率である。

相関係数が0.8を超えると
予測「個人 p の方が特性Bが強い」
の外れる確率も5人に1人以下になる。

「個人 p の方が特性Aが強い」とき、
言明「個人 p の方が特性Bが強い」
の成り立つ確率は、
個人間で特性Aの差が大きい正の値であるほど高くなる
と考えられる。

差 $A_p - A_q$ の関数として確率

$$P(B_p - B_q > 0 | A_p - A_q)$$

を描くと図3のようになる。

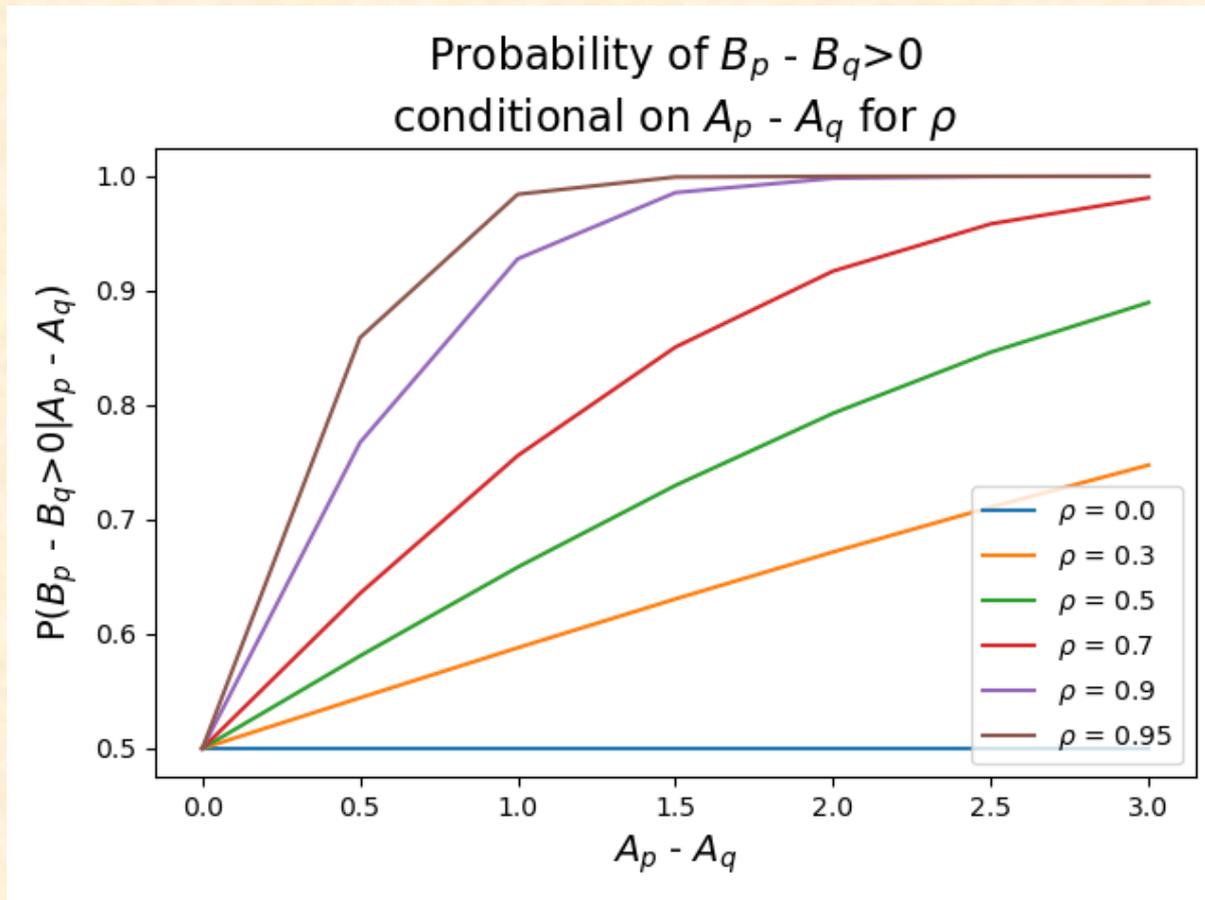


図3 差 $A_p - A_q$ と確率 $P(B_p - B_q > 0 | A_p - A_q)$

相関係数 ρ が0の場合は、
 $A_p - A_q > 0$ の値に関わらず
予測「個人 p の方が特性Bが強い」
が成り立つ確率は0.5である。

相関係数が0.3であれば、
差 $A_p - A_q$ の値が2.0のとき
3人に1人ぐらいの割合で言明は成り立たない。

相関係数が0.9であれば、
差 $A_p - A_q$ の値が2以上のとき
言明が成り立つ確率は1.0であると言える。

相関係数によって2変量（特性）の関係の強さがどう変わるかを図示するために、相関係数 ρ の2変量正規分布を表す次式の等高線図を描いてみた（図4）。

$$\begin{bmatrix} A \\ B \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

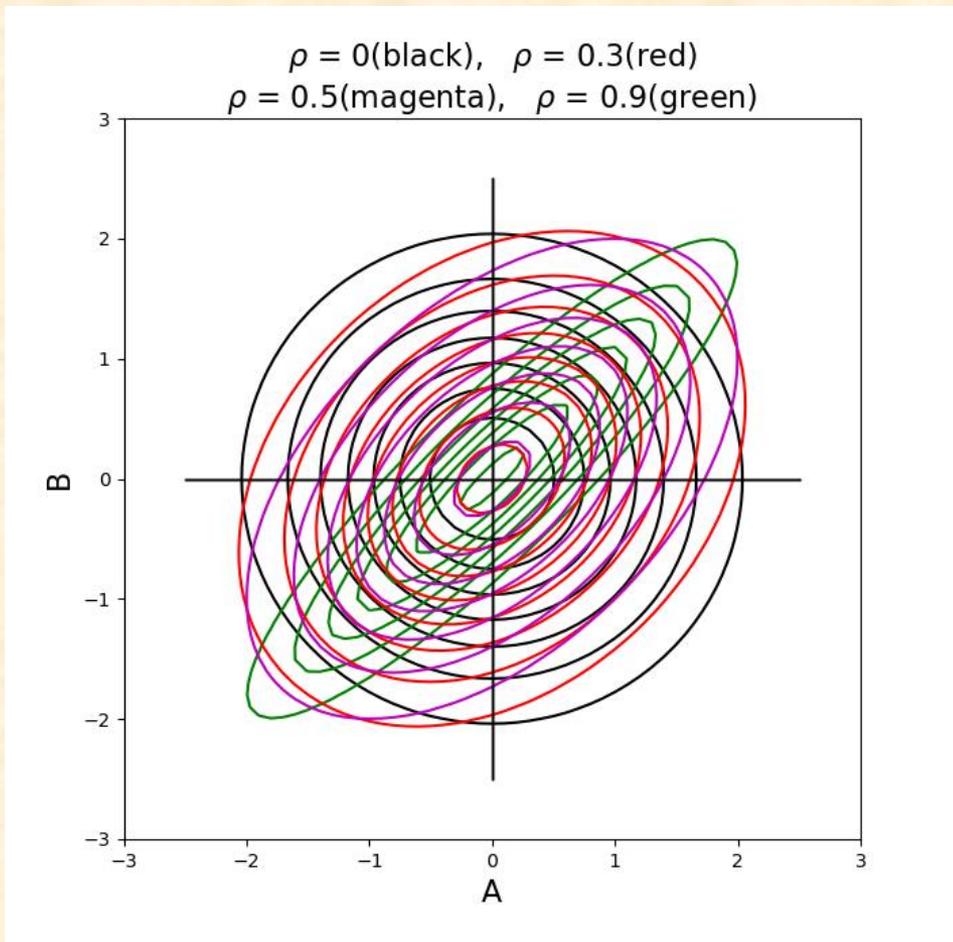


図4 2変量正分布の等高線図

無相関 $\rho = 0$ のときは、等高線は円である。

相関係数 $\rho = 0.3$ のときも円に近い。

相関係数 $\rho = 0.9$ になれば分布は

$$\text{直線 } Y = X$$

の周りに集まっているので、

X の値と Y の値がともに同じように変化する確率が高いと期待できる。

まとめ

「一方の特性が強いと他方の特性も強い」という判断（予測）は、相関の有無に基づくとき悉無的に行うことは誤りである。

相関係数は確率的関係を表すので、それに基づく特性間の関係判断の正しさも確率的である。相関係数**0.3**は中ほどの関係の強さを表すとされているが、このときの「一方の特性が強いと他方の特性も強い」という予測の成立確率は**0.5**に近く（図1）、これは当てずっぽうの予測と大差ない値である。

特性値の強さを考慮した場合は、一方の特性値が強くなるほど予測「他方も強い」が成立する確率も高くなるが、相関係数が**0.3**の場合、 1σ 程度の強さでは3人に1人の割合で成り立たない（図2）。

個人間の比較の場合も、同様である。

予測の成立が確率的である場合、その確率を示すことは、科学的アセスメントの1つの要件であろう。