

因果関係を可変とする回帰分析(その1)

—提案と一般論—

日大生産工 〇篠原正明
情報システム研究所 篠原健

1. はじめに

従来の回帰分析を適用する理論枠組では、1つの被説明変量データ（例えば、 X_j ）を残りの説明変量データ（ $X_i, i=1, \dots, N, i \neq j$ ）でいかに説明するかという立場から、最小二乗法、誤差和最小法など様々なアプローチが採用されて来た。ここで、1つの被説明変量はその説明変量群により効果的に説明できるという暗黙の仮定が存在する。効果的に説明できる関係が存在するという事は、被説明変量が結果事象で、説明変量が原因事象に対応する大きな可能性を示唆しており、本論文ではこのような立場をとる。ところで、回帰分析適用時には、変量間の因果関係（正確には、説明—被説明関係と言うべきところであるが、本論文では因果関係と呼ぶ）は必ずしも確定しているわけではなく、もっともらしい因果関係を仮定したもとの回帰分析を適用する場面も多くある（と聞く）。すなわち、変量間の因果関係が必ずしも明確でない場面において、変量間の適切な因果関係も含めて最適な回帰分析を行う必要が生じる。そこで、本論文では、事象 $E_i (i = 1, \dots, N)$ に関する観測データ X_i が与えられているが、事象間の因果関係が不明な場合の多変量データ解析法として**可変因果回帰分析 (VCRA: Variable-Causality Regression Analysis)** を提案する。

2. 可変因果回帰分析の一般論

N 個の事象間には事象対の選び方の数は $C(N, 2) = N! / (N-2)! 2!$ 存在する。例えば、 $N=3$ では $C(3, 2) = 3$ (図 1)、 $N=4$ では $C(4, 2) = 6$ (図 2) である。

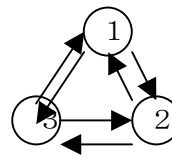


図 1 3 事象間因果関係の離接枝表現

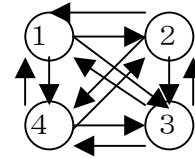


図 2 4 事象間因果関係の離接枝表現

事象対 (i, j) 間の因果関係としては、事象 i が原因で事象 j が結果 ($i \rightarrow j$)、事象 j が原因で事象 i が結果 ($i \leftarrow j$)、事象 i, j 間に因果関係なし (i と j 間に枝無し) の 3 通りを考えると、 N 事象の場合には、 $3^{C(N, 2)}$ 個の因果グラフが存在しうる。図 3 に、 $N=3$ で、27 通りの因果グラフを示す。ここで、例えば、(012) は第 1 項 0 に対応して枝 (1, 2) 間が無因果、第 2 項 1 に対応して枝 (1, 3) 間が $1 \rightarrow 3$ の順因果、第 3 項 3 に対応して枝 (2, 3) 間が $3 \rightarrow 2$ の逆因果を表現する。

3. おわりに

複数事象間の因果関係が不明で、対応する変量データの標本値が与えられている場合に、特定事象対 (i, j) 間の因果関係として、無因果、順因果 ($i \rightarrow j$)、逆因果 ($j \rightarrow i$) の 3 つを考え、それらの組み合わせにより因果グラフの 3 値離接枝表現を提案した。事象数 $N=3$ の場合でも、様々な因果関係が存在しうるし、 $N > 3$ の場合には、その種類の多様性はさらに増大する。また、本アプローチの前提として、「本来の因果関係を想定した場合には、モデルとデータの適合度が高まる」を仮定しているが、因果関係の多様性のもとにおける適切な適合度指標の選定は今後の大きな課題の 1 つである。

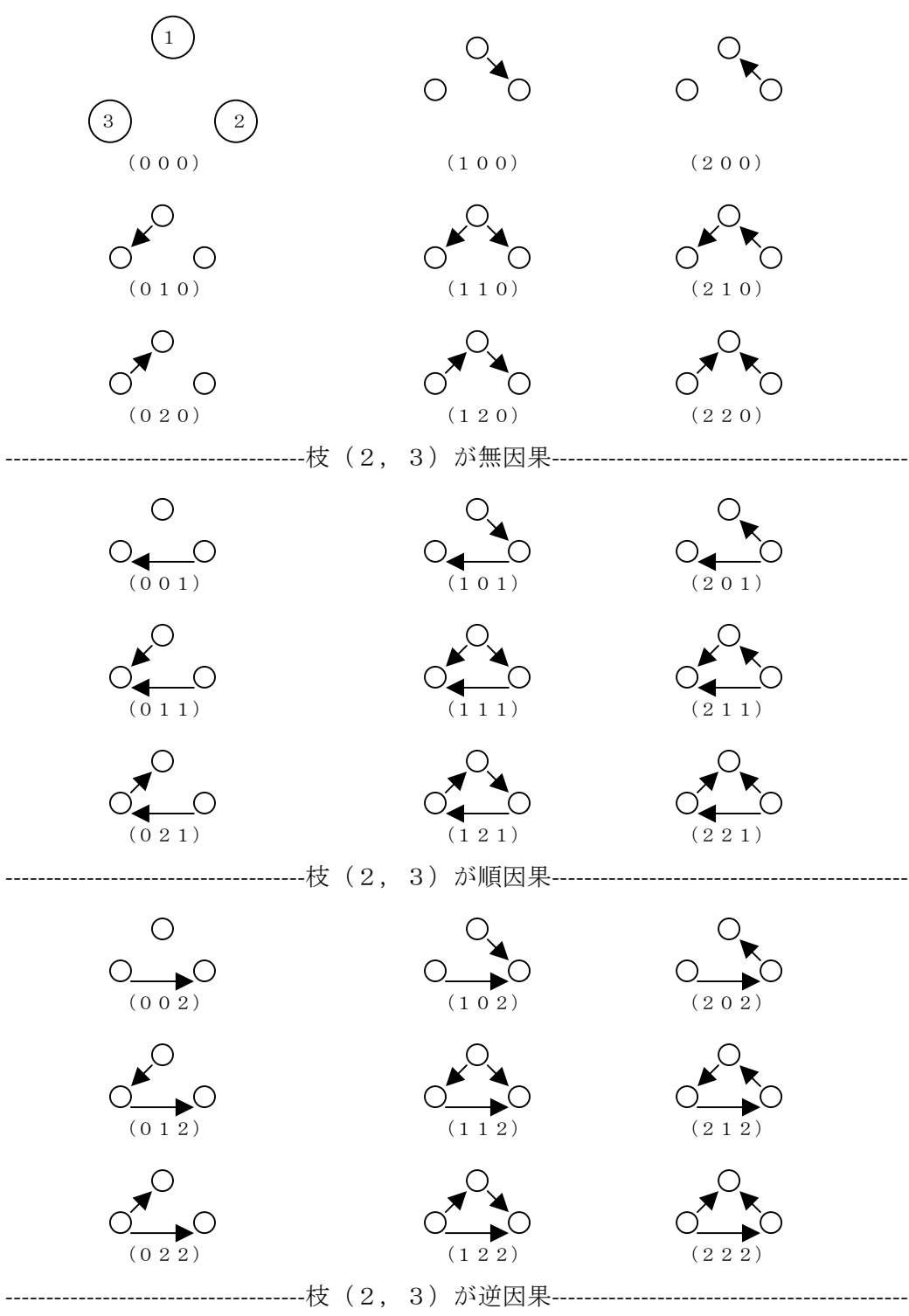


図 3 事象数 N=3 の場合の全ての因果グラフ (27通り)