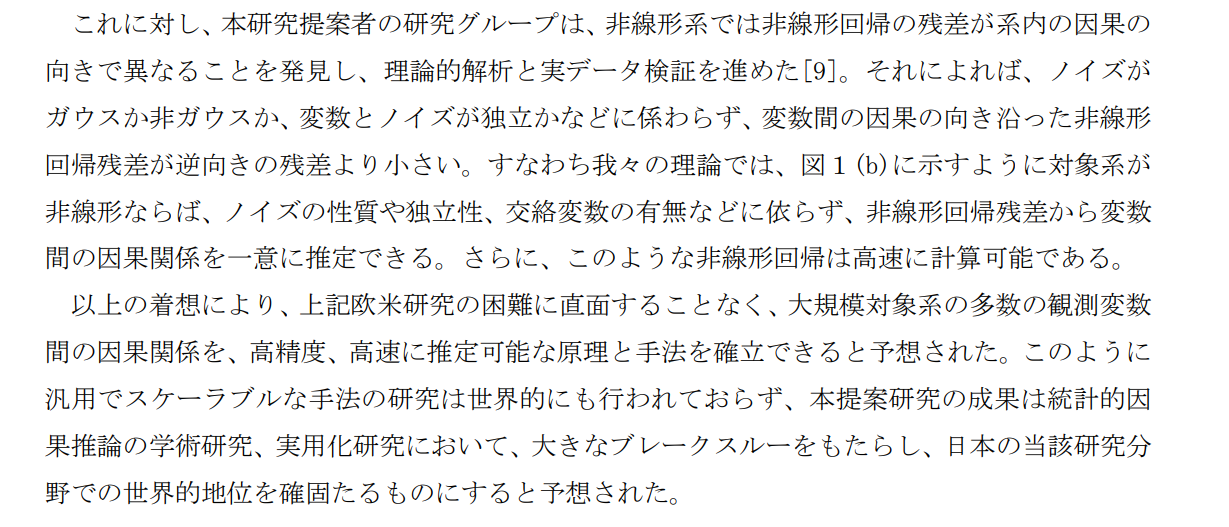
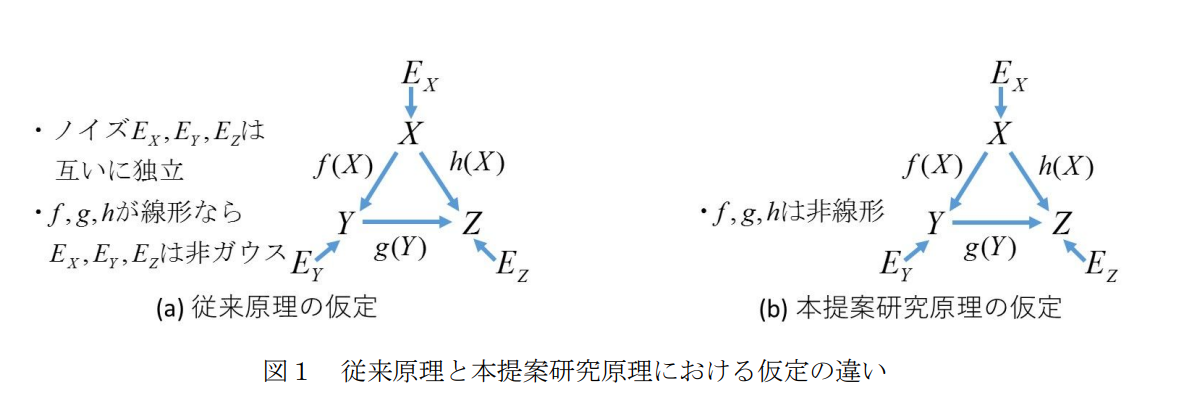
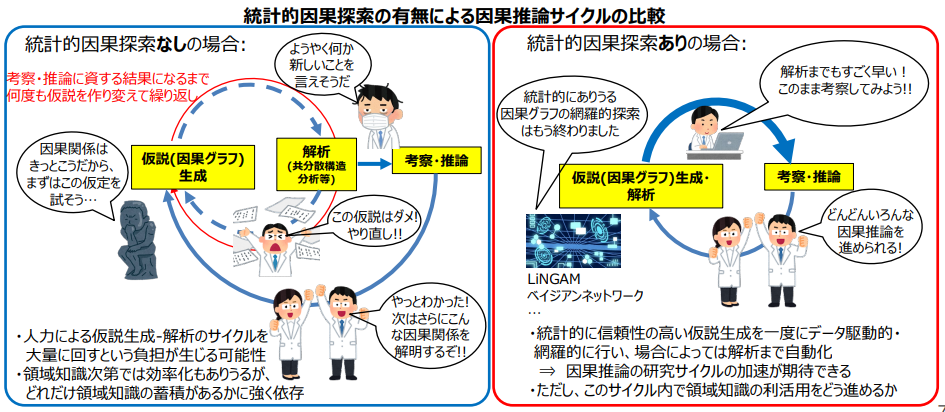
<https://kaken.nii.ac.jp/ja/file/KAKENHI-PROJECT-17K00305/17K00305seika.pdf>







<https://www.nistep.go.jp/wp/wp-content/uploads/09_review20220218.pdf>

**潜在共通変数（未観測の交絡因子）がある場合のLiNGAM**

LiNGAMモデル

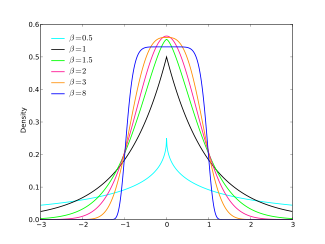
は残差が互いに独立であることを前提としている。

これは潜在共通変数がデータ内に無い事を前提としている。潜在共通変数がある場合のモデルは

で、潜在共通変数を何らかの形で決める必要がある点とそもそも潜在共通変数の個数Lも不明なためそれをも決める必要がある。そもそも潜在共通変数は観測されていないのでを何らかの方法で推定できれば良い。を一つの切片としてしてしまえば潜在共通原因の個数Lを推定する必要は無い。

その代わりにの分布自体が未知量という問題になる。

既知または観測していないだけの潜在共通変数は限られているが現実世界では潜在共通変数は観測されていないため無数に存在すると考えられる。その場合の分布は中心極限定理により正規分布に近い分布になると考えられる。そこでの分布は不明だが正規分布に近い分布を含む一般化ガウス分布を設定しての分布を決めるパラメータ()が未知とする事で出来る。



残差が互いに独立であることを前提としているため適切な()が見つかれば残差が互いに独立になると考えられる。しかし現実にはそのような正確な分布を求められないため残差が互いに独立に近いように()を見つけるような計算を行う事は可能です。ただし、単に残差が互いに独立に近いようにするのは難しくなく殆ど無関係な状態になるように()を決める事が出来てしまう。そこで線形回帰としてデータを可能な限りフィット出来る出来なければならないという制約が必要になる。

**提案手法**

最適化問題へ帰着させて初めから全体の因果構造を求める。この時の最適化条件とパラメータは

1. **回帰モデルとし正しい（残差が最小）**

**（２）残差と残差間の相互情報量がなるべく最小化するようにの分布パラメータを求める**。

**（３）因果関係の事前知識があれば事前知識によるペナルティーを科す**

回帰モデルとして精度を上げる一方で各変数の残差が独立になるようにの分布パラメータを求める。

ペナルティーは既知の知識による因果関係は維持する為に導入している。

最適化を行う損失関数は「拡大Tchebyshev・スカラー化関数」を使う。

**簡易的な推定方法**

初期値として可能性のある因果構造を洗い出す。つまり多くの作成する必要がある。

そのためにまず一旦の分布を決める。つまりパラメータ()を設定しての分布を決める。

観測変数をで補正　　すると補正前と異なる因果構造を示すが得られる。ただし、得られた因果構造が同一形状な物も含まれるが係数の値は同一になる事は殆ど無い。

パラメータ()の範囲を設定して各パラメータ()から因果構造の候補を生成する事が出来る。

これによって

から残差が得られるのでが最小な結果となる因果構造を見つければよい。

**非線形への拡張**

Post-Nonlinear Causal Model

はの親（原因）となる変数。提案した線形モデルと同様な考え方を踏襲して

は観測された親変数以外の共通潜在原因です。線形モデルと違い切片といった形で一つの分布に押し込むことは出来ないため複数の未知の異なる分布が複数あると仮定する。

の分布は不明だが正規分布に近い分布を含む一般化ガウス分布を設定しての分布を決めるパラメータ()が未知とする。

非線形回帰モデルとして精度を上げる一方で各変数の残差が独立になるようにの分布パラメータを求める。ペナルティーは既知の知識による因果関係は維持する為に導入している。

最適化を行う損失関数は「拡大Tchebyshev・スカラー化関数」を使う。

**簡易的な推定方法**

初期値として可能性のある因果構造を洗い出す。つまり多くの作成する必要がある。

そのためにまず一旦の分布を決める。つまりパラメータ()を設定しての分布を決める。

観測変数をで補正　　すると補正前と異なる因果構造を示すが得られる。ただし、得られた因果構造が同一形状な物も含まれるが係数の値は同一になる事は殆ど無い。

非巡回の因果関係であれば以下の行列を定義できる。

行方向が子（結果）で列方向が親（原因）として与える事が出来る。

パラメータ()の範囲を設定して各パラメータ()から因果構造の候補を生成する事が出来る。

これによってニューラルネットでと、を学習させる事が出来る。

そして学習結果から

から残差が得られるのでが最小な結果となる因果構造を見つければよい。

**LiNGAM in the presence of latent common variables (unobserved confounders)**

LiNGAM model

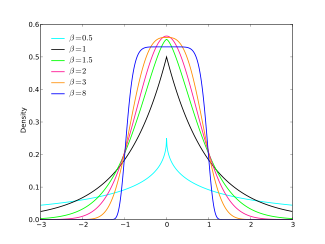
assumes that the residuals are independent of each other.

This assumes that there is no latent common variable in the data. The model when there is a latent common variable is

で、latent common variableThe number of latent common variablesis not known, so it is also necessary to determine the number L. To begin with, the latent common variablehas not been observed, so。as one intercept, then the number of potential common causes L There is no need to estimate the number of latent common causes L.

Instead, the distribution of itself becomes a matter of unknown quantity.

There are a limited number of latent common variables that are known or just unobserved, but in the real world, latent common variables are considered to be innumerable because they are unobserved. In such a case, the distribution ofis considered to be close to a normal distribution according to the central limit theorem. Therefore, it is possible to set up a generalized Gaussian distribution that includes a distribution close to the normal distribution, although the distribution of is unknown, by assuming that the parameters (β, ρ) that determine the distribution of μ are unknown.



It is assumed that the residuals are independent of each other if an appropriate (β,ρ) is found. However, in reality, it is not possible to find such an exact distribution, so it is possible to perform calculations such that (β,ρ) is found so that the residuals ε\_i are close to independent of each other. However, it is not difficult to simply make the residuals close to independent of each other, but it is possible to determine (β,ρ) so that they are almost unrelated. Therefore, we need to make a restriction that we must be able to fit the data as much as possible as a linear regression.

**Proposed Method**

The entire causal structure is obtained from the beginning by attributing it to an optimization problem. The optimization conditions and parameters are

(1) The regression model is assumed to be correct (residuals are minimized).

(2) Find the distribution parameter of so that the amount of mutual information between residuals and residuals is minimized as much as possible.

(3) If there is prior knowledge of causality, penaltyF is imposed due to prior knowledge.

The distribution parameter of is obtained so that the residuals of each variable are independent while increasing the accuracy of the regression model.

The penalty is introduced to maintain the causal relationship based on known knowledge.

**Simplified estimation method**

Identify possible causal structures as initial values. In other words, create many .

First, determine the distribution of once. In other words, set the parameters (β,ρ) to determine the distribution of .

Correcting the observed variables by 　, we obtain that show a causal structure different from that before the correction. However, the obtained causal structure may have the same shape, but the coefficient values are rarely the same.

It is possible to set the range of parameters (β, ρ) and generate a candidate causal structure from each parameter (β, ρ).

This allows

Since the residuals are obtained from , it is sufficient to find the causal structure that results in the minimum of .