时序ICP

□ 定义不变集S：

$$(\mathbf{Y},\mathbf{X}) = (Y_t, X_t)_{t\in\{1,\ldots,n\}} \in \mathbb{R}^{n\times(d+1)}$$

A set $S \subseteq \{1,\ldots,d\}$ is called invariant with respect to $(\mathbf{Y},\mathbf{X})$ if there exist parameters $\mu \in \mathbb{R}$, $\beta \in (\mathbb{R}\setminus\{0\})^{|S|\times 1}$ and $\sigma \in \mathbb{R}_{>0}$ such that

(a) $\forall t \in \{1,\ldots,n\}: \quad Y_t = \mu + X_t^S\beta + \varepsilon_t$ *and* $\varepsilon_t \perp\!\!\!\perp X_t^S$,

(b) $\varepsilon_1,\ldots,\varepsilon_n \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$.

$$H_{0,S}: \text{ S is an invariant set with respect to } (\mathbf{Y},\mathbf{X}).$$

$$\Downarrow$$

$$H_{0,S}: \begin{cases} \exists\beta \in (\mathbb{R}\setminus\{0\})^{|S|},\ \sigma\in(0,\infty): \\ \mathbf{Y} = \mathbf{X}^S\beta + \boldsymbol{\varepsilon},\ \text{with } \boldsymbol{\varepsilon} \perp\!\!\!\perp \mathbf{X}^S \text{ and } \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\sigma^2\mathbf{Id}), \end{cases}$$

- **目标：** 基于观测数据$(\mathbf{Y},\mathbf{X})$，估计 S^*
- **方法：** 对所有不变集S求交 $\tilde{S} := \bigcap_{\substack{S\subseteq\{1,\ldots,d\}: \\ H_{0,S} \text{ is true}}} S \subseteq S^*$

[Pfister N, Bühlmann P, Peters J. Invariant causal prediction for sequential data[J]. Journal of the American Statistical Association, 2019, 114(527): 1264-1276.]

时序ICP及其扩展作用

东南大学 SOUTHEAST UNIVERSITY

- 在**无环境**（不需要不同环境先验信息），**时序数据**（非独立同分布），**无时滞效应**情况下推断 $\forall t \in \{1,\ldots,n\}: \quad Y_t = \mu + X_t^S \beta + \varepsilon_t \ and \ \varepsilon_t \perp\!\!\!\perp X_t^S$；**包含ICP为特例。**

- 推断多元**线性**时间序列中的**瞬时+时滞因果效应**，优于格兰杰因果。

$(\mathbf{Y}, \mathbf{X}) = (Y_t, X_t)_{t \in \{1,\ldots,n\}} \in \mathbb{R}^{n \times (d+1)}$, $S^* \subseteq \{1,\ldots,d\}$, $\beta = (\beta_1, \ldots, \beta_{|S^*|})^\top \in (\mathbb{R} \setminus \{0\})^{|S^*| \times 1}$

$B_k \in \mathbb{R}^{(d+1)\times 1}$ for $k \in \{1,\ldots,p\}$, satisfying for all $t \in \{p+1,\ldots,n\}$ that

$$Y_t = X_t^{S^*} \beta + \sum_{k=1}^{p} (Y_{t-k}, X_{t-k}) B_k + \varepsilon_t,$$

结构向量自回归

- 推断多元**非线性**时序中的**瞬时因果效应**，包含**非线性ICP**为特例。

$$Y_t = f(X_t^{S^*}, \varepsilon_t)$$

[Pfister N, Bühlmann P, Peters J. Invariant causal prediction for sequential data[J]. Journal of the American Statistical Association, 2019, 114(527): 1264-1276.]

瞬时因果效应

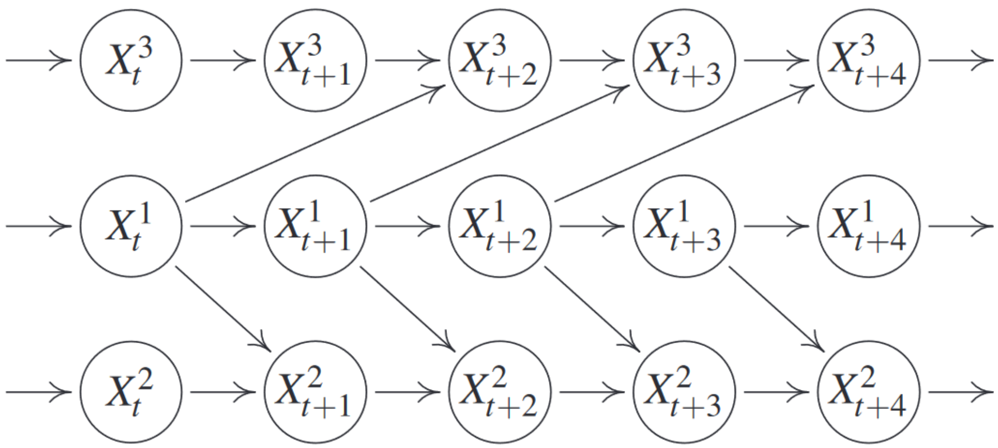
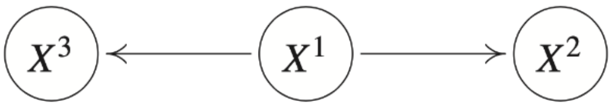


Figure 10.1: Example of a time series with no instantaneous effects.

Summary graph：



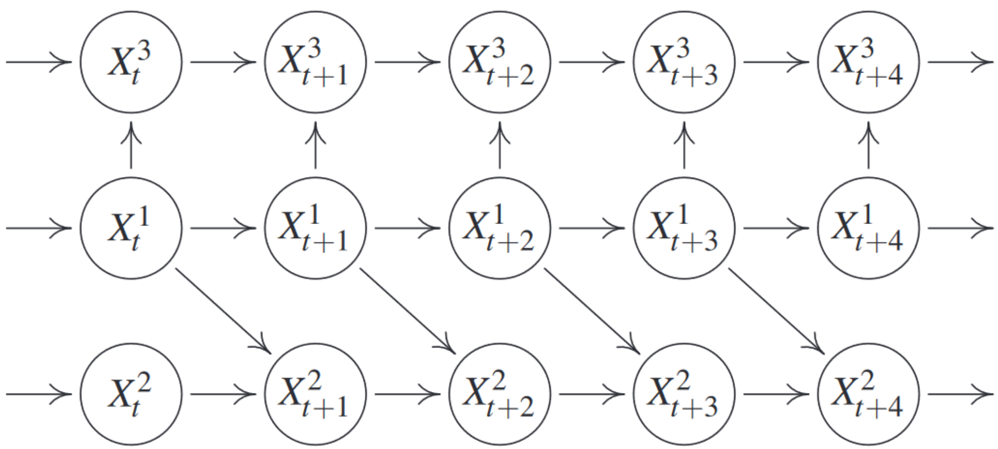
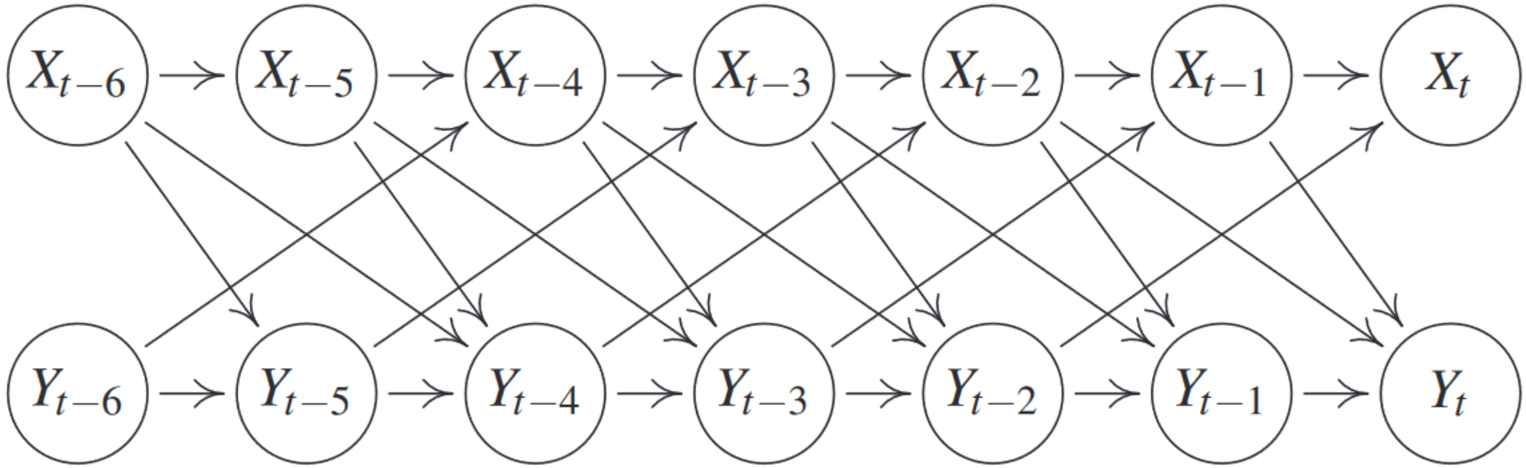


Figure 10.2: Example of a time series with instantaneous effects.

[Peters J, Janzing D, Schölkopf B. Elements of causal inference: foundations and learning algorithms[M]. The MIT Press, 2017.]

格兰杰因果



$$X \text{ Granger-causes } Y \quad :\Longleftrightarrow \quad Y_t \not\perp\!\!\!\perp X_{\text{past}(t)} \,|\, Y_{\text{past}(t)}.$$

$$Y_t = \sum_{i=1}^{q} a_i Y_{t-i} + N_t$$

$$Y_t = \sum_{i=1}^{q} a_i Y_{t-i} + \sum_{i=1}^{q} b_i X_{t-i} + \tilde{N}_t,$$

$\Longleftrightarrow$ **含X项时回归所得残差的方差明显更小，则称 X Granger-causes Y。**

[Peters J, Janzing D, Schölkopf B. Elements of causal inference: foundations and learning algorithms[M]. The MIT Press, 2017.]

结构向量自回归

□ 向量自回归：$\{x_i(t)|i=1,\ldots,n;t=1,\ldots,T\}$

在**向量自回归模型**（VAR，Vector Autoregressive Model）中，同一样本期间内的 n 个变量（内生变量）可以作为它们过去值的线性函数， k 阶的 VAR 模型可以表示如下：

$$\boldsymbol{x}(t)=\sum_{\tau=1}^{k}\boldsymbol{B}_\tau\boldsymbol{x}(t-\tau)+\boldsymbol{e}(t)$$

其中， k 为模型的时间间隔数量， $\boldsymbol{B}_\tau,\tau=1,\ldots,k$ 为每一个时间间隔对应的权重矩阵，$\boldsymbol{e}(t)$ 为未被观测的噪声。

□ 结构方程模型：

在**结构方程模型中**（SEM，structural equation models）中，认为因果效应是瞬时的，没有时间间隔，观测变量被简单地建模为其他观测变量的线性函数：

$$\boldsymbol{x}=\boldsymbol{B}\boldsymbol{x}+\boldsymbol{e}$$

其中 $\boldsymbol{B}$ 为瞬时的权重矩阵， $\boldsymbol{e}$ 时服从某种分布的噪声向量

$$\boldsymbol{x}(t)=\sum_{\tau=0}^{k}\boldsymbol{B}_\tau\boldsymbol{x}(t-\tau)+\boldsymbol{e}(t)$$

结构向量自回归

[Hyvärinen A, Zhang K, Shimizu S, et al. Estimation of a structural vector autoregression model using non-gaussianity[J]. Journal of Machine Learning Research, 2010, 11(5).]

时序ICP实验-线性


```r
set.seed(1)
# environment 1
na <- 140
X1a <- 0.3*rnorm(na)
X3a <- X1a + 0.2*rnorm(na)
Ya <- -.7*X1a + .6*X3a + 0.1*rnorm(na)
X2a <- -0.5*Ya + 0.5*X3a + 0.1*rnorm(na)

# environment 2
nb <- 80
X1b <- 0.3*rnorm(nb)
X3b <- 0.5*rnorm(nb)
Yb <- -.7*X1b + .6*X3b + 0.1*rnorm(nb)
X2b <- -0.5*Yb + 0.5*X3b + 0.1*rnorm(nb)

# combine environments
X1 <- c(X1a,X1b)
X2 <- c(X2a,X2b)
X3 <- c(X3a,X3b)
Y <- c(Ya,Yb)
Xmatrix <- cbind(X1, X2, X3)

summary(lm(Y~Xmatrix))

# apply seqICP to the same setting
seqICP.result <- seqICP(X = Xmatrix, Y, par.test = list(grid = seq(0, na + nb, (na + nb)
summary(seqICP.result) # seqICP is able to infer that X1 and X3 are causes of Y
```

```
> summary(lm(Y~Xmatrix))

Call:
lm(formula = Y ~ Xmatrix)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.205831 -0.061317 -0.001113  0.057515  0.266640 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.001799   0.005980   0.301    0.764    
XmatrixX1   -0.583158   0.027397 -21.285  < 2e-16 ***
XmatrixX2   -0.379482   0.047765  -7.945 1.06e-13 ***
XmatrixX3    0.687121   0.018082  38.000  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Invariant Linear Causal Regression at level 0.05
Variables X1, X3 show a significant causal effect

          coefficient lower bound upper bound  p-value
intercept         0.0    -0.05900      0.0179       NA
X1               -0.7    -0.75200     -0.5292     0.02 *
X2                0.0     0.00000      0.0000     0.32
X3                0.6     0.57000      0.7228     0.02 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

时序ICP实验-非线性

```r
set.seed(2)

# environment 1
na <- 120
X1a <- 0.3*rnorm(na)
X3a <- X1a + 0.2*rnorm(na)
Ya <- 2*X1a^2 + 0.6*sin(X3a) + 0.1*rnorm(na)
X2a <- -0.5*Ya + 0.5*X3a + 0.1*rnorm(na)

# environment 2
nb <- 80
X1b <- 2*rnorm(nb)
X3b <- rnorm(nb)
Yb <- 2*X1b^2 + 0.6*sin(X3b) + 0.1*rnorm(nb)
X2b <- -0.5*Yb + 0.8*rnorm(nb)

# combine environments
X1 <- c(X1a,X1b)
X2 <- c(X2a,X2b)
X3 <- c(X3a,X3b)
Y <- c(Ya,Yb)
Xmatrix <- cbind(X1, X2, X3)

# use GAM as regression function
GAM <- function(X,Y){
  d <- ncol(X)
  if(d>1){
    formula <- "Y~1"
    names <- c("Y")
    for(i in 1:(d-1)){
      formula <- paste(formula,"+s(X",toString(i),")",sep="")
      names <- c(names,paste("X",toString(i),sep=""))
    }
    data <- data.frame(cbind(Y,X[,-1,drop=FALSE]))
    colnames(data) <- names
    fit <- fitted.values(mgcv::gam(as.formula(formula),data=data))
  } else{
    fit <- rep(mean(Y),nrow(X))
  }
  return(fit)
}

# Y follows the same structural assignment in both environments
# a and b (cf. the lines Ya <- ... and Yb <- ...).
# The direct causes of Y are X1 and X3.
# A GAM model fit considers X1, X2 and X3 as significant.
# All these variables are helpful for the prediction of Y.
summary(mgcv::gam(Y~s(X1)+s(X2)+s(X3)))

# apply seqICP to the same setting
seqICPnl.result <- seqICPnl(X = Xmatrix, Y, test="block.variance",
                   par.test = list(grid = seq(0, na + nb, (na + nb)/10), comp
                            alpha = 0.05, B =100), regression.fun = GA

summary(seqICPnl.result)
```

```
Formula:
Y ~ s(X1) + s(X2) + s(X3)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.430047   0.006571     522   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
        edf Ref.df        F  p-value    
s(X1) 9.000  9.000 1690.766  < 2e-16 ***
s(X2) 6.302  7.612    5.957 1.95e-06 ***
s(X3) 6.396  7.577  227.789  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =      1   Deviance explained =   100%
GCV = 0.0097406  Scale est. = 0.0086352  n = 200
```

```
> summary(seqICPnl.result)

 Non-linear Invariant Causal Regression at level 0.05
 Variables X1, X3 show a significant causal effect
   p-value   
X1 0.009901 **
X2 0.386139   
X3 0.009901 **
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```