



# Analyses de trajectoires par classes latentes : nouveaux développements

*Club SAS de Québec  
7 juin 2022*

**Denis Talbot, PhD  
Professeur titulaire  
Université Laval**



# Plan

- ▶ Comment quantifier l'incertitude par rapport au modèle choisi ?
- ▶ Comment estimer l'effet des trajectoires sur une réponse ?



## Quantifier l'incertitude – Motivation

- ▶ Plusieurs décisions sur la spécification du modèle prises à partir des données
- ▶ Est-ce que d'autres données auraient menées à choisir le même modèle ?
- ▶ Les indices d'ajustement actuels ne répondent pas à cette question

# BIC

- ▶ À partir des BIC des modèles, on peut approximer la probabilité a posteriori des modèles par (Clyde, 2003)

$$P(M_j|D) \approx \frac{\exp(-BIC_j/2)}{\sum_k \exp(-BIC_k/2)}$$

- ▶ ATTENTION : PROC TRAJ rapporte

$$BIC^* = BIC / -2$$

Il faut donc calculer

$$P(M_j|D) \approx \frac{\exp(BIC_j^*)}{\sum_k \exp(BIC_k^*)}$$

- ▶ Au besoin, on peut ajouter une même constante à tous les BIC\*

# Exemple

```
/* 1 groupe */  
PROC TRAJ DATA=OPPOSITN;  
  ID ID; VAR O1-O7; INDEP T1-T7;  
  MODEL CNORM; MAX 10; NGROUPS 1; ORDER 2;  
RUN;  
  
/* 2 groupe */  
PROC TRAJ DATA=OPPOSITN;  
  ID ID; VAR O1-O7; INDEP T1-T7;  
  MODEL CNORM; MAX 10; NGROUPS 2; ORDER 2 2;  
RUN;  
  
...  
  
/*6 groupes*/  
PROC TRAJ DATA=OPPOSITN;  
  ID ID; VAR O1-O7; INDEP T1-T7;  
  MODEL CNORM; MAX 10; NGROUPS 6; ORDER 2 2 2 2 2 2;  
RUN;
```



# Exemple

Group membership

1	(%)	68.24299	5.43977	12.545	0.0000
2	(%)	31.75701	5.43977	5.838	0.0000

BIC= -1640.80 (N= 901)    BIC= -1633.30 (N= 138)    AIC= -1621.59    ll= -1613.59

Nb groupes	BIC*	Prob. Post.
1	-1701.24	0%
2	-1640.80	25%
3	-1639.70	75%
4	-1650.76	0%
5	-1659.53	0%
6	-1670.55	0%



# Bootstrap

Nous avons proposé une approche basée sur le bootstrap (Mésidor et al, 2022).

1. Rééchantillonner avec remise  $n$  observations parmi  $n$
2. Choisir le meilleur modèle selon le BIC pour le nouvel échantillon
3. Répéter 1-2 un nombre  $B$  fois
4. Évaluer la variabilité dans le modèle choisi



## Exemple – rééchantillonnage

```
PROC SURVEYSELECT  
  DATA = OPPOSITN  
  SEED = 94171  
  SAMPRATE = 1  
  METHOD = URS  
  OUTHITS  
  OUT = bootsamp  
  REP = 100;  
RUN;
```





## Exemple – Choisir le meilleur modèle

```
%MACRO boot_traj;  
%DO I = 1 %TO 100;  
    DATA bootsampi;  
        SET bootsamp;  
        WHERE replicate = &I;  
    RUN;  
  
    PROC TRAJ DATA=bootsampi OUTEST=OE1 OUT=OF;  
        ID ID; VAR O1-O7; INDEP T1-T7;  
        MODEL CNORM; MAX 10; NGROUPS 1; ORDER 2;  
    RUN;  
  
    ...  
  
    DATA OE;  
        MERGE OE1(rename = (_BIC1_ = BIC1) where = (_TYPE_ = "PARMS"))  
              ...  
              OE6(rename = (_BIC1_ = BIC6) where = (_TYPE_ = "PARMS"));  
        maxBIC = max(BIC1, BIC2, BIC3, BIC4, BIC5, BIC6);  
        IF BIC1 = maxBIC THEN best = 1;  
        ...  
        rep = &I;  
        KEEP best rep;  
    RUN;
```



## Exemple – Choisir le meilleur modèle (suite)

```
%IF &I = 1 %THEN %DO;  
    DATA results_boot;  
        SET OE;  
    RUN;  
%END; %ELSE %DO;  
    DATA results_boot;  
        SET results_boot OE;  
    RUN;  
%END;  
%END;  
%MEND boot_traj;  
%boot_traj;
```



## Exemple – variabilité

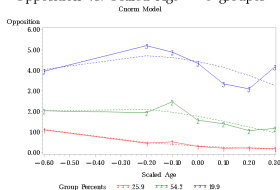
Nb groupes	Prop réplifications
1	0%
2	13%
3	66%
4	15%
5	4%
6	2%

- ▶ Résultats différents vs le BIC
- ▶ Nos résultats suggèrent que le bootstrap est plus fiable
- ▶ Rapporter les résultats de différents modèles si l'incertitude est grande

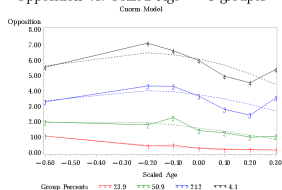


# Exemple – Résultats

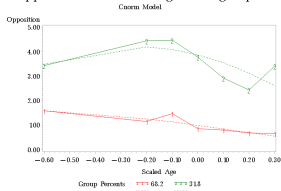
Opposition vs. Scaled Age – 3 groupes



Opposition vs. Scaled Age – 4 groupes



Opposition vs. Scaled Age – 2 groupes





# Estimer l'effet des trajectoires – Motivation

Les analyses de trajectoires peuvent être utilisées pour différentes raisons



# Estimer l'effet des trajectoires – Motivation

Les analyses de trajectoires peuvent être utilisées pour différentes raisons

- ▶ Analyses descriptives
- ▶ Modélisation



# Estimer l'effet des trajectoires – Motivation

Les analyses de trajectoires peuvent être utilisées pour différentes raisons

- ▶ Analyses descriptives
- ▶ Modélisation
  - ▶ Comme variable réponse



# Estimer l'effet des trajectoires – Motivation

Les analyses de trajectoires peuvent être utilisées pour différentes raisons

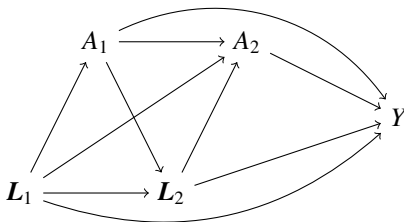
- ▶ Analyses descriptives
- ▶ Modélisation
  - ▶ Comme variable réponse
  - ▶ **Comme variable explicative**





# Défi : Confusion dépendante du temps

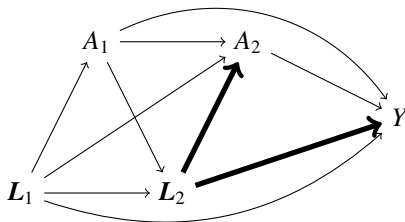
Exemple : on s'intéresse à l'effet de trajectoires de  $A$  sur  $Y$





# Défi : Confusion dépendante du temps

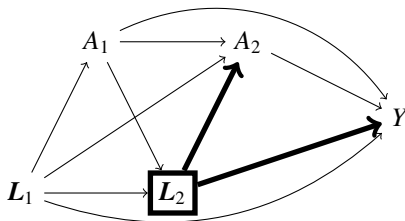
Exemple : on s'intéresse à l'effet de trajectoires de  $A$  sur  $Y$





# Défi : Confusion dépendante du temps

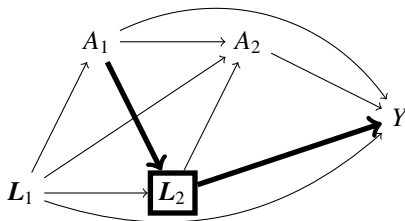
Exemple : on s'intéresse à l'effet de trajectoires de  $A$  sur  $Y$





# Défi : Confusion dépendante du temps

Exemple : on s'intéresse à l'effet de trajectoires de  $A$  sur  $Y$





# Modèles structurels marginaux

Les modèles structurels marginaux (MSM) permettent de régler ce problème

- ▶ Chaque temps d'exposition  $A_t$  est ajusté pour ses propres confondants
- ▶ La pondération par probabilité inverse d'exposition est un estimateur relativement simple à implanter



# LCGA-MSM

Nous avons développé une approche qui combine l'analyse de trajectoires (LCGA) et les modèles structurels marginaux (Diop et al, 2021)

1. Effectuer une analyse de trajectoires pour former des groupes  $G$  de trajectoires
2. Pour chaque sujet, estimer les poids

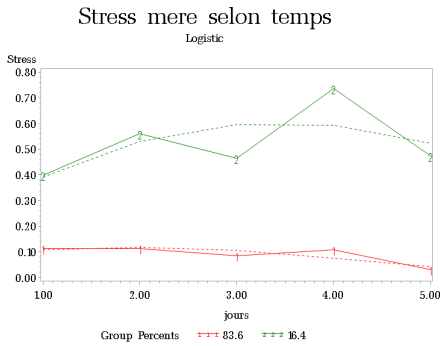
$$w_i = \prod_{t=1}^T \frac{P(A_t = a_{i,t} | \bar{A}_{i,t-1})}{P(A_t = a_{i,t} | \bar{L}_{i,t}, \bar{A}_{i,t-1})}$$

3. Effectuer un modèle  $g(\mathbb{E}[Y|G]) = \beta_0 + \beta_1 G_1 + \beta_2 G_2 + \dots + \beta_3 G_{J-1}$  sur les données pondérées
4. Utiliser un estimateur sandwich de la variance pour les inférences



# Exemple – 1. Analyse de trajectoires

```
PROC TRAJ DATA = ds OUTPLOT=OP OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;  
  ID id;  
  VAR s2-s6; INDEP t1-t5;  
  MODEL LOGIT; NGROUPS 2; ORDER 2 2;  
RUN;  
%TRAJPLOT(OP,OS,'Stress mere selon temps','Logistic','Stress','jours')
```



## Exemple – 2. Estimer les poids

```
PROC LOGISTIC DATA = ds DESCENDING;  
  MODEL s2 = s1 i1 married employed csex;  
  OUTPUT OUT = ds_denom2 P = p_denom2;  
RUN;  
PROC LOGISTIC DATA = ds DESCENDING;  
  MODEL s3 = s2 i2 married employed csex;  
  OUTPUT OUT = ds_denom3 P = p_denom3;  
RUN;  
...  
PROC LOGISTIC DATA = ds DESCENDING;  
  MODEL s2 = s1;  
  OUTPUT OUT = ds_num2 P = p_num2;  
RUN;  
PROC LOGISTIC DATA = ds DESCENDING;  
  MODEL s3 = s2;  
  OUTPUT OUT = ds_num3 P = p_num3;  
RUN;  
...
```





## Exemple – 2. Estimer les poids (suite)

```
/* Calcul des poids */  
DATA poids;  
    MERGE ds_num2-ds_num6 ds_denom2-ds_denom6;  
    BY id;  
    w2 = s2*(p_num2/p_denom2) +  
        (1 - s2)*(1 - p_num2)/(1 - p_denom2);  
    w3 = s3*(p_num3/p_denom3) +  
        (1 - s3)*(1 - p_num3)/(1 - p_denom3);  
    w4 = s4*(p_num4/p_denom4) +  
        (1 - s4)*(1 - p_num4)/(1 - p_denom4);  
    w5 = s5*(p_num5/p_denom5) +  
        (1 - s5)*(1 - p_num5)/(1 - p_denom5);  
    w6 = s6*(p_num6/p_denom6) +  
        (1 - s6)*(1 - p_num6)/(1 - p_denom6);  
    w = w2*w3*w4*w5*w6;  
    KEEP id w i;  
RUN;  
  
PROC MEANS DATA = poids MIN Q1 MEAN MEDIAN Q3 MAX;  
    VAR w;  
RUN;
```



## Exemple – 3+4. Modélisation réponse

```
DATA ds_msm;  
    MERGE poids of;  
    BY id;  
RUN;  
  
/* Brut */  
PROC GENMOD DATA = ds_msm;  
    CLASS id group;  
    MODEL i = group / DIST = normal;  
    REPEATED SUBJECT = id;  
RUN;  
  
/* Ajuste */  
PROC GENMOD DATA = ds_msm;  
    CLASS id group;  
    MODEL i = group / DIST = normal;  
    REPEATED SUBJECT = id;  
    WEIGHT w;  
RUN;
```



## Exemple – Résultats

Modèle	Coefficient	Estimé	IC 95%
Brut	Ordonnée	0.20	( 0.09, 0.30)
	Groupe 1	-0.07	(-0.18, 0.05)
Ajusté	Ordonnée	0.13	( 0.09, 0.30)
	Groupe 1	-0.00	(-0.12, 0.12)



# Conclusion – Forces et limites

Avantages des analyses de trajectoires :

- ▶ Approche visuelle permettant de décrire les tendances temporelles
- ▶ Relativement simple à implanter avec SAS

Limites des analyses de trajectoires :

- ▶ Existe-t-il vraiment des groupes de trajectoires ?
- ▶ Hypothèse d'indépendance locale



# Conclusion – Bootstrap

- ▶ Permet de quantifier l'incertitude par rapport au modèle choisi
- ▶ Un peu plus difficile et long à implanter que le BIC



## Conclusion – LCGA-MSM

- ▶ Procédure en trois étapes relativement simples
- ▶ Approche théoriquement justifiée pour estimer l'effet des trajectoires
- ▶ Résultats conditionnels aux trajectoires trouvées, sans supposer qu'elles sont réelles



# Références

- ▶ Clyde M. Subjective and objective Bayesian statistics, 2nd ed. S. James Press, New Jersey: Wiley-Interscience, 2003
- ▶ Mésidor, M., Sirois, C., Simard, M., & Talbot, D. (2022). A bootstrap approach for validating the number of groups identified by latent class growth models. arXiv preprint arXiv:2205.07631.
- ▶ Diop, A., Sirois, C., Guertin, J. R., & Talbot, D. (2021). Marginal structural models with Latent Class Growth Modeling of Treatment Trajectories. arXiv preprint arXiv:2105.12720