

# Méthodes d'appariement

## Université d'été en Sciences Sociales

### Tany Vao 2022 - Madagascar

Florent Bédecarrats, Jeanne de Montalembert,  
Marc Bouvier, Marin Ferry, Kenneth HOUNGBEDJI

Université de Toliara, Octobre 2022

## Principe des méthodes d'appariement

Est ce que les Aires Protégées augmentent le couvert forestier?

On veut connaitre :

$$E[CF1|AP = 1] - E[CF0|AP = 0]$$

avec  $CF1$  le couvert forestier s'il y a une AP et  $CF0$  le couvert forestier s'il n'y a pas d'AP **au même moment**.

Autrement dit,  $E[Y1|T = 1] - E[Y0|T = 0]$

On n'observe pas simultanément les deux: quel **contrefactuel** ?

- Différents espaces ?

## Principe des méthodes d'appariement

Espace	Couvert forestier
Makay	$CF_M = M + AP$
Autre	$CF_A = A$

Note:  $M$  représente l'effet fixe *Makay* et  $A$  l'effet fixe de l'*Autre*

Effet causal:  $CF_M - CF_A = M + \underbrace{AP}_{\text{effet}} - A$

$$\rightarrow M - A$$

$M - A$  correspond aux différences de couvert forestier sans AP

⇒ **biais**

# Principe des méthodes d'appariement

Le principe de la méthode d'appariement est de minimiser la différence entre  $M$  et  $A$  *i.e* de trouver un espace qui ressemble le plus possible à l'espace qui est protégé, son jumeau.

On va chercher le "jumeau" grâce à des caractéristiques que l'on peut mesurer *i.e* des **caractéristiques observables**  $X$  et on va comparer les résultats entre ces espaces qui se ressemblent !

## Principe des méthodes d'appariement

Autrement dit, conditionnellement à des variables observables  $X$ , l'affectation au traitement est indépendante des résultats potentiels *i.e*

$$(Y_0, Y_1) \perp T | X$$

$\perp$  orthogonalité, indépendance *i.e* ne dépend pas,  $T|X$  traitement étant donné les caractéristiques observables

# Hypothèse centrales

## Hypothèses sous-jacentes:

- L'assignation à un programme est aléatoire lorsque les deux unités ont les mêmes caractéristiques observables  
⇒ **Indépendance conditionnelle (aux caractéristiques observables)**
- Il doit y avoir suffisamment d'unités traitées qui ressemblent à des unités contrôles  
⇒ **Support commun**

# Hypothèse d'indépendance conditionnelle

## L'hypothèse d'indépendance conditionnelle:

- L'assignation dans un programme de conservation d'un espace ayant les mêmes caractéristiques  $X$  est indépendante des résultats potentiels avec et sans le traitement.

→ Conditionnellement aux caractéristiques observables  $X$ , les résultats moyens des unités qui ne bénéficient pas du programme peuvent servir de **contrefactuel** des résultats moyens des unités qui bénéficient du programme en l'absence de ce dernier.

⇒ L'indépendance conditionnelle permet de ne plus avoir de **problème de sélection** *i.e* d'obtenir des **estimateurs sans biais** de l'effet du traitement (conditionnel aux observables).

## Hypothèse d'indépendance conditionnelle

Assouplissement de l'hypothèse d'indépendance conditionnelle lorsque l'on s'intéresse à l'effet du traitement sur les unités traitées (ATT):

- Hypothèse d'indépendance conditionnelle est moins forte *i.e*

$$\left( Y(0) | X \perp T | X \right)$$

→ Conditionnellement aux caractéristiques observables  $X$ , le résultat des unités non-traitées (= qui n'ont pas bénéficié de la mesure  $T = 0$ ) représente un **contrefactuel non biaisé** du résultat potentiel moyens des unités pour lesquelles  $T = 1$  en l'absence de traitement  $T$ .



## Hypothèse d'indépendance conditionnelle

Implication de l'hypothèse d'indépendance conditionnelle (CIA)

⇒ En dehors des observables ( $X$ ), il n'existe pas d'autres caractéristiques qui influencent **à la fois** les résultats potentiels ( $Y$ ) et l'adoption du programme ( $T$ ).

Risque important que l'hypothèse ne soit pas respectée.

→ Pour pouvoir comparer une unité traitée avec une unité non traitée identique, il faut qu'une **unité non traitée identique existe** !

→ Si, pour certaines caractéristiques  $X$ , les unités sont assurées d'être traitées alors il ne sera **pas possible** de trouver des unités non traitées comparables !

# Hypothèse d'existence d'un support commun

## L'hypothèse d'existence d'un support commun ("overlap")

- Avoir des unités traitées et non traitées **qui se ressemblent** donc avoir un **chevauchement suffisant** entre les unités traitées et les unités non traitées pour trouver des appariement.

→ Pour chaque valeur de  $X$ , on suppose qu'il existe des unités qui ont le traitement et d'autres qu'y ne l'ont pas.

# Le principe

## Méthodes d'appariement

On va chercher une unité  $\Rightarrow$  Pour évaluer l'effet moyen sur les traités (ATE), cela revient à **comparer** le résultat  $Y_1$  de chaque unité traitée avec le résultat  $Y_0$  de l'unité non traitée ayant exactement les mêmes caractéristiques observables

ieurs variantes dans la manière de construire le contrefactuel de chaque unité et donc la manière de trouver un "jumeau" varient en fonction de la technique utilisée:

- Voisin le plus proche
- Voisins les plus proches
- Score de propension...

# Le principe

## Méthodes d'appariement

- L'idée est de comparer les résultats ( $Y_1$ ) de chaque unité avec les résultats ( $Y_0$ ) son "jumeau"

⇒ Pour évaluer l'effet moyen sur les traités (ATE), cela revient à **comparer** le résultat  $Y_1$  de chaque unité traitée avec le résultat  $Y_0$  de l'unité non traitée ayant exactement les mêmes caractéristiques observables

ieurs variantes dans la manière de construire le contrefactuel de chaque unité et donc la manière de trouver un "jumeau" varie en fonction de la technique utilisée:

- Voisin le plus proche
- Voisins les plus proches
- Score de propension...

# Voisin le plus proche

## Voisin le plus proche

- La méthode la plus simple est de comparer chaque unité avec son "jumeau" *i.e* des unités qui ont les **caractéristiques observables**  $X$  les plus proches
- Il est rare de trouver des unités non traitées avec des caractéristiques identiques à celles des unités traitées.  
⇒ On choisit donc son "**plus proche voisin**"  $i$
- Cette notion demande de définir une **métrique de la distance** ou de la **ressemblance entre les unités**

# Les métriques

Il y a plusieurs manières de calculer la distance entre deux unités :

- **Distance euclidienne:** la distance entre deux unités est la somme des distance de toutes les covariables *i.e* les caractéristiques observables  $X$ .
- **Distance de Mahalanobis:** accord un poids différents aux caractéristiques observables  $X$ .

# Variantes du plus proche voisin

## Les plus proches voisins

- Plus proches voisins:
  - On apparie avec un nombre fixe  $m$  de plus proches voisins et le résultat potentiel contrefactuel est donc la moyenne des résultats de ces  $m$  voisins
  - On exclut les couples trop éloignés *i.e* on n'utilise pas les unités traités pour lesquels on ne peut pas trouve un (ou  $m$ ) jumeau.x à moins d'une certaine distance  $d$  à fixer.

# Mécanique de l'appariement

L'appariement peut être fait:

- **Sans remise:** une unité du groupe de contrôle ne peut être appariée qu'une fois avec une unité du groupe de traitement
- **Avec remise:** on utilise l'ensemble de l'échantillon à chaque fois ce qui autorise des appariements avec la même unité

Contraintes du matching sans remise:

- un large échantillon de contrôle
- l'estimation pourra être sensible à l'ordre dans lequel l'appariement est effectué



## Limites du plus proche voisin

L'appariement est assez simple à mettre en oeuvre, son principe est intuitif et ne demande pas de choix de paramètres.

Toutefois, on ne contrôle pas la **qualité** de l'appariement, la notion du plus proche voisin est par nature relative.

→ Certains plus proches voisins pourront être en fait éloignés. Or la méthode du plus proche voisin traite de la même manière des couples proches et moins proches.

Apparier avec un seul individu prive de l'information apportée par tous les autres ce qui **réduit a priori la précision** de l'estimation.

→ Par exemple, certains bénéficiaires peuvent avoir plusieurs jumeaux très proches: on peut considérer qu'il est dommage d'en choisir un seul (plus ou moins arbitrairement).

## Choix de la taille de la fenêtre

La fenêtre  $d$  mesure la taille du voisinage en dehors duquel les poids sont très faible.

Plus la fenêtre  $d$  est petite, plus l'estimation du contrefactuel d'une unité traité ne prendra en compte que les unités non traitées dont les caractéristiques observables sont très proches des siennes.

Il n'existe pas vraiment de règle établie pour le choix de cette fenêtre.

En pratique, le choix de la fenêtre est fait de manière *ad-hoc*

## Précision vs biais

Chaque méthode a des avantages et des inconvénients.

L'opposition entre la plus simple (voisin le plus proche) et la plus élaborée (noyau) reflète le dilemme classique entre biais et variance:

- Ne pas utiliser l'ensemble de l'information disponible réduit la précision
- Les estimations par fonction à noyau sont toujours plus précises, même sur de petits échantillons (Frolich, 2004), mais on peut craindre qu'elles augmentent aussi les risques de mauvais appariements et donc les biais.

En pratique, on recommande de tester la sensibilité des résultats à la méthode utilisée.

## Combien de variables de contrôle ?

Le matching exact est souvent compliqué à mettre en oeuvre.

Pour que l'hypothèse CIA soit vérifiée, on souhaite utiliser le maximum d'information et donc apparier sur de très nombreuses variables. Mais plus on intègre des variables, plus il s'avèrera difficile de trouver un voisin proche.

A distance finie, les estimateurs sont d'autant plus biaisés que le nombre de variables de conditionnement  $X$  est élevé (d'autant plus lorsque les variables de conditionnement sont continues).

→ L'appariement sur le score de propension (PSM) est pragmatique et désirable.

## Appariement sur le score de propension

Si l'hypothèse CIA est vérifiée, alors les résultats potentiels sont indépendants de l'adhésion au traitement conditionnellement au score de propension (Rosenbaum et Rubin, 1983).

Soit le score de propension  $\pi(X_i) \equiv P(T_i = 1|X_i)$  et  $0 < \pi(X_i) < 1$  i.e la probabilité d'être traité conditionnellement aux observables, on a:

$$\left( Y(0), Y(1) \right) \perp T|X \Rightarrow \left( Y(0), Y(1) \right) \perp T|\pi(X_i)$$

Cette propriété diminue le nombre de dimension des comparaisons puisque l'on ramène le nombre de variables de conditionnement à une seule et unique variable, qui serait un résumé univarié de l'ensemble des covariables.

Cependant, le PSM n'est parfois non connu et il est donc nécessaire de l'estimer.

## Estimation du score de propension

Pour tenir compte de la nature bornée du PSM, il est d'usage de le modéliser par une forme logistique (ou un probit):

Même s'il est coutume d'estimer le PSM par un modèle logit ou probit, il convient de se rappeler que la forme fonctionnelle du PSM est souvent inconnue.

## Restriction au support commun

Les méthodes présentées sous l'hypothèse d'indépendance conditionnelle reposent sur l'hypothèse d'un support commun.

Celle-ci signifie qu'il existe à la fois des unités traitées et des non traitées ayant des valeurs des observables identiques *i.e* qui partagent le même PSM.

Si ce n'est pas le cas, il sera impossible de trouver pour chaque unités traitées des non traitées comparables.

→ Il est important de vérifier que le **support commun** *i.e* la zone sur laquelle il existe des unités traitées et non traitées qui partagent le même score de propension, est suffisamment large.

## Test visuel du support commun

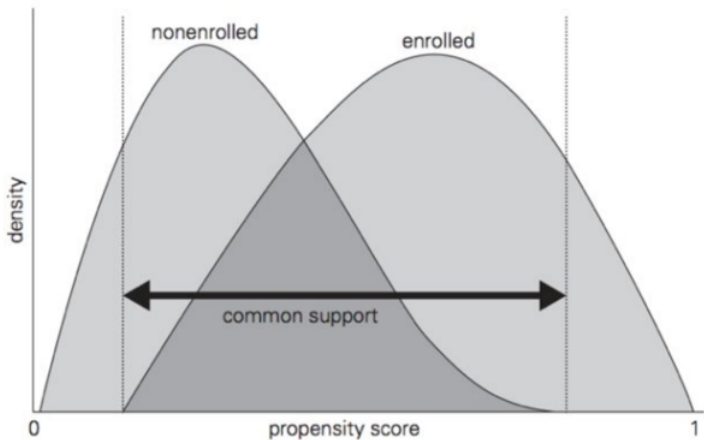
On peut représenter le support commun à partir de la distribution des scores de propension des unités traitées et non-traitées.

En pratique, on représente les histogrammes de probabilité estimée d'être traité pour les unités bénéficiaires et des non bénéficiaires du traitement étudié.

On peut ensuite vérifier l'étendue des valeurs de  $\pi(X_i)$  pour lesquelles on observe un nombre suffisant d'unité traitée et non-traitée *i.e* la largeur du support commun.



## Test visuel du support commun



**Figure 1:** Exemple de support commun.

## Enjeux liés au support commun

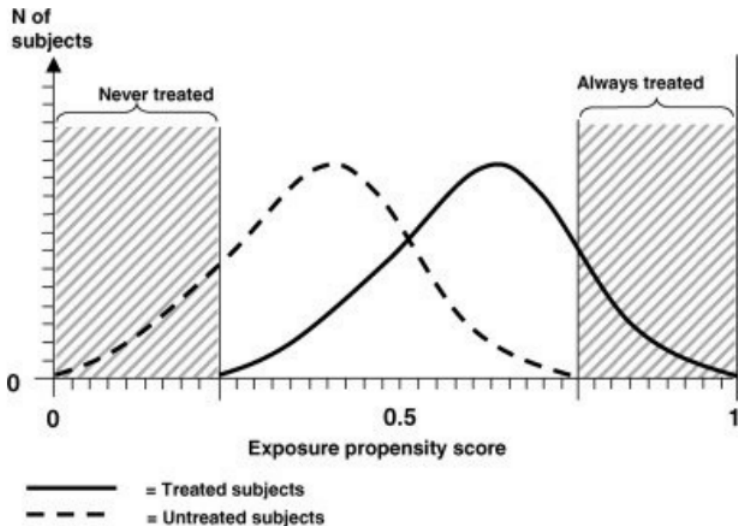
L'hypothèse d'indépendance conditionnelle suppose qu'on compare des personnes avec des caractéristiques observables identiques.

- Si elles n'existent pas pour certaines valeurs des observables, il est incorrect d'utiliser ces unités pour l'observation
- Si on ne se limite pas au support commun, les estimations pourront être biaisées

En pratique, plusieurs méthodes ont été proposées pour se restreindre au support commun.

Ces méthodes changent la nature de ce qui est estimé puisqu'on estime un impact sur une partie de la population dont les observables sont telles qu'on observe un recouvrement pour les deux sous échantillons.

## Enjeux liés au support commun



**Figure 2:** Exemple de support commun et validité externe

# Imposer un support commun

## Méthode du min/max

- Pour estimer l'ATT, Dehejia et Wahba (1999) éliminent les unités non traitées ayant un score inférieur au minimum observé sur les unités traités *i.e* "*never treated*"
- Pour estimer l'ATE, il faut également éliminer les unités traitées ayant un score supérieur au maximum observé pour les unités non traitées *i.e* "*always treated*"

## Quelles variables de conditionnement/contrôle ?

L'hypothèse d'indépendance conditionnelle demande d'observer les caractéristiques qui déterminent l'allocation du traitement.

En pratique, il doit s'agir de variables qui peuvent avoir un impact sur la variable d'intérêt (e.g déforestation) et sur le choix de bénéficiaire du dispositif (e.g Aire Protégée).

Deux points de vigilance:

- Il ne faut pas utiliser de variables mesurées après la mise en place du dispositif et qui peuvent également affectée celui-ci → problème d'endogénéité
- Il faut qu'il existe une part d'aléa ou d'exogénéité dans la sélection des unités traitées sinon il n'y a plus de support commun puisque la participation au programme est entièrement déterminé par des caractéristiques des unités

## Estimateur

L'effet moyen estimé sur les traités correspond à la moyenne de ces effets estimés:

$$\hat{E}[Y_i(1) - Y_i(0) | T_i = 1] = \frac{1}{n_1} \sum_{i \in E_1} (y_i(1) - \hat{y}_i(0))$$

où  $n_1$  est le nombre d'unités traitées

$E_1$  est l'ensemble des unités traitées

$y_i(1)$  le résultat observé de l'individu  $i$

$\hat{y}_i(0)$  le résultat du plus proche voisin non traité de  $i$

## Résumé

Pour mettre en oeuvre une méthode d'appariement, il faut:

- Identifier les unités traitées et les non-traitées
- Observer les caractéristiques qui déterminent le traitement pour chaque unité
- Choisir une méthode d'estimation: plus proche(s) voisin(s), PSM ?
- Dans le cas d'un appariement sur le score de propension, il faut l'estimer.
- Estimer les effets (ATT et/ou ATE)

## Résumé

La crédibilité des estimateurs estimés repose sur l'hypothèse très forte d'indépendance conditionnelle.

- Tiens aux observables **disponibles et utilisées** pour la comparaison
- Si on dispose de peu d'information dans nos données sur les choix entre les variables, il ne sera pas possible d'éliminer le **biais de sélection**