



THE WORLD BANK



L'Inférence causale

Felipe Barrera

Human Development
Network

Middle East and North Africa
Region

Spanish Impact Evaluation
Fund

Motivation

- ❑ Les questions de recherche qui motivent la plupart des études dans les sciences de la santé sont causales de par leur nature.
- ❑ Par exemple :
 - Quelle est l'efficacité d'un certain médicament sur une population ciblée ?
 - Quelle fraction de décès par une certaine maladie donnée aurait pu être évitée par un traitement donné ou une politique donnée?

Motivation

- Les questions empiriques les plus difficiles en économie impliquent également des relations d'effet causal :
 - La décentralisation scolaire améliore-t-elle la qualité des écoles ?
 - Une année de plus d'éducation apporte-t-elle un revenu plus élevé ?
 - La distribution d'allocations familiales conditionnées (conditional cash transfers) améliore-t-elle la santé des enfants ?

Motivation

L'intérêt à ces questions est motivé par :

- Des préoccupations de politique
 - ▣ Les programmes publics réduisent-ils la pauvreté ?

- Des considérations théoriques

- Des problèmes auxquels se heurtent les décideurs individuels

Intuition du problème : un exemple hypothétique



- Une allocation pour les femmes enceintes, payée sous certaines conditions:
 - Assistance à un contrôle médical périodique
 - Assistance à un programme de nutrition
 - Assistance à des réunions périodiques sur les risques du tabagisme et de l'alcool

- La question clé: ce programme a-t-il un impact sur le poids des enfants à la naissance ?

Intuition du problème

- Données hypothétiques : comparons le poids à la naissance d'enfants nés dans une ville où se trouve le programme, à celui des enfants nés dans une ville voisine où il n'y a pas ce programme

Poids à la naissance, grammes (moyenne)			
	Familles vivant dans une ville avec le programme	Familles vivant dans une ville voisine sans programme	Différence
Avant le programme	?	?	
Après le programme	3,250	3,100	150

Intuition du problème

- Données hypothétiques : comparons le poids à la naissance d'enfants nés dans une ville où se trouve le programme, à celui des enfants nés dans une ville voisine où il n'y a pas ce programme

Poids à la naissance, grammes (moyenne)				
	Familles vivant dans une ville avec le programme	Familles vivant dans une ville voisine sans programme	Différence	Double Différence
Avant le programme	3,025	2,840	185	
Après le programme	3,250	3,100	150	-35

Intuition du problème

- Données hypothétiques : Assignons les familles au programme par tirage aléatoire

Poids à la naissance, grammes (moyenne)				
	Familles vivant dans une ville avec le programme	Familles vivant dans une ville voisine sans programme	Différence	Double Différence
Avant le programme	3,025	2,840	185	
Après le programme	3,250	3,100	150	-35
	Familles avec programme TIRAGE ALEATOIRE	Familles sans programme TIRAGE ALEATOIRE	Différence	Double Différence
Avant le programme	3,028	3,028	0	
Après le programme	3,250	3,105	145	145
Hypothèse : les différences sont significatives				

Analyse causale

- Le but de l'analyse statistique standard, personnifiée par la probabilité et d'autres techniques d'estimation, est d'inférer certains paramètres d'une distribution à partir d'échantillons tirés de cette distribution.

- Avec l'aide de ces paramètres, on peut :
 1. Inférer une association entre différentes variables,
 2. Estimer la probabilité d'événements passés et futurs,
 3. Mettre à jour l'estimation de la probabilité de certains événements lorsque de nouvelles mesures deviennent disponibles.

Analyse causale

- Ces tâches sont bien gérées par l'analyse statistique standard pour autant que les conditions expérimentales restent les mêmes.

- L'analyse causale va plus loin :
 - Elles vise à inférer des aspects du processus de génération de données.

 - A l'aide de ces aspects, on peut déduire non seulement la probabilité d'événements dans des conditions statiques, mais aussi la dynamique d'événements dans des *conditions changeantes*.

Analyse causale

- Cette capacité inclut :
 1. Prédire les effets de certaines interventions
 2. Prédire les effets de changements spontanés
 3. Identifier les causes d'événements notifiés

- Cette distinction implique que les concepts causaux et associationnels ne se mélangent pas.

Analyse causale

Le mot *cause* n'est pas dans le vocabulaire de la théorie de probabilité standard.

- Toute théorie de probabilité nous permet de dire que deux événements sont mutuellement corrélés, ou dépendants – signifiant que si nous trouvons u , nous pouvons nous attendre à rencontrer l'autre.
- Les scientifiques recherchant des explications causales pour des phénomènes complexes ou des justifications pour les décisions de politique doivent par conséquent compléter le langage de la probabilité avec un vocabulaire de causalité.

Analyse causale

- Deux langages pour la causalité ont été proposés :
 1. Les équations structurelles de modélisation (ESM) (Haavelmo 1943).
 2. Le modèle potentiel de résultats de Neyman-Rubin (RCM) (Neyman, 1923; Rubin, 1974).

Le modèle causal de Rubin

- Définissons la population par U . Chaque unité dans U est dénotée par u .
- A chaque $u \in U$, est associée une valeur $Y(u)$ de la variable d'intérêt Y , que nous appelons la variable de réponse.
- Soit A une seconde variable définie sur U . Nous appelons A un attribut des unités dans U .

- ▣ La notion clé est le *potentiel* pour exposer ou ne pas exposer chaque unité à l'action d'une cause :

- ▣ Chaque unité doit être potentiellement exposable à chacune des causes.

- ▣ Donc, Rubin **prend** la position selon laquelle les causes ne sont que ces choses qui pourraient être des traitements dans des expériences hypothétiques.

- ▣ Un attribut ne peut pas être une cause dans une expérience, parce que la notion d'*exposabilité potentielle* ne s'applique pas à lui.

- Pour simplifier, nous assumons qu'il y a seulement deux causes ou niveaux de traitement.

- Soit D une variable qui indique la cause à laquelle chaque unité dans U est exposée :

$$D = \begin{cases} t & \text{si l'unité } u \text{ est exposée au traitement} \\ c & \text{si l'unité } u \text{ est exposée au témoin} \end{cases}$$

Dans une étude sous contrôle, D est construit par l'expérimentateur. Dans une étude non contrôlée, il est déterminé par des facteurs au-delà du contrôle de l'expérimentateur.

- Les valeurs de Y sont potentiellement affectées par la cause particulière, t ou c , à laquelle l'unité est exposée.

- Donc, nous avons besoin de deux variables de réponse :

$$Y_t(u), Y_c(u)$$

- Soit $Y_t(u)$ la valeur de la réponse observée si l'unité u fusse exposée à t
- Soit $Y_c(u)$ la valeur de la réponse observée si l'unité u fusse exposée à c .

- Laissons également D être exprimé comme une variable binaire (indicatrice):

$$D = 1 \text{ if } D = t \text{ and } D = 0 \text{ if } D = c$$

- Donc, le résultat de chaque individu u peut être écrit comme étant :

$$Y(u) = D Y_1 + (1 - D) Y_0$$

- **Définition** : Pour chaque unité u , le traitement $\{c'est\text{-à-dire } D_u = 1\}$ lieu de $D_u = 0\}$ cause l'effet suivant:

$$\delta_u = Y_1(u) - Y_0(u)$$

- Cette définition d'un effet causal assume que le statut de traitement d'un individu n'affecte pas les résultats potentiels d'autres individus.
- **Le problème fondamental de l'inférence causale** : Il est impossible d'observer la valeur de $Y_1(u)$ et $Y_0(u)$ sur la même unité u . Par conséquent, il est impossible d'observer l'effet de t sur u .
- Un autre moyen d'exprimer le problème est de dire que nous ne pouvons pas inférer l'effet de traitement parce que nous n'avons pas la preuve *contrefactuelle*. C'est-à-dire que pour les individus traités, nous ne savons pas ce qui se serait passé en l'absence de traitement.

□ Etant donné que l'effet causal pour une certaine unité u ne peut pas être observé, nous cherchons à identifier l'**effet causal moyen** pour la population dans son ensemble ou pour certaines sous-populations.

□ L'**effet de traitement moyen** « ATE » de t (par rapport à c) sur U (ou toute sous-population) est donné par:

$$\begin{aligned}
 ATE &= E [Y_1(u) - Y_0(u)] \\
 &= E [Y_1(u)] - E [Y_0(u)] \\
 &= \bar{\delta} = \bar{Y}_1 - \bar{Y}_0 \quad (1)
 \end{aligned}$$

- ▣ Vu qu'il est impossible d'observer l'effet causal de t sur une
 unité spécifique u, nous utilisons une solution statistique qui est
 d'estimer l'effet causal *moyen* de t sur une population d'unités.

- ▣ Bien que $E(Y_1)$ et $E(Y_0)$ ne puissent tous deux être calculés, ils
 peuvent être estimés.

- ▣ La plupart des méthodes économétriques tentent de construire
 des estimations cohérentes de

\bar{Y}_1 et \bar{Y}_0
 à partir de données observationnelles

□ Soit le suivant un estimateur simple de ATE :

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0] \quad (2)$$

- Notons que l'équation (1) est définie pour l'ensemble de la population, alors que l'équation (2) représente un estimateur à être évalué sur un échantillon tiré de cette population

□ Soit π la proportion de la population qui serait affectée au groupe de bénéficiaires.

□ En décomposant ATE, nous obtenons :

$$\begin{aligned}
 \bar{\delta} &= \pi \bar{\delta}_{\{D=1\}} + (1 - \pi) \bar{\delta}_{\{D=0\}} \\
 &= \pi \left[(\bar{Y}_1 - \bar{Y}_0) | D = 1 \right] + (1 - \pi) \left[(\bar{Y}_1 - \bar{Y}_0) | D = 0 \right] \\
 &= \left[\pi [\bar{Y}_1 | D = 1] + (1 - \pi) [\bar{Y}_1 | D = 0] \right] + \\
 &\quad \left[\pi [\bar{Y}_0 | D = 1] + (1 - \pi) [\bar{Y}_0 | D = 0] \right] \\
 &= \bar{Y}_1 - \bar{Y}_0
 \end{aligned}$$

- Si nous assumons que :

~~$$[\bar{Y}_1 | D = 1] = [\bar{Y}_1 | D = 0] \text{ et } [\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$~~

- Nous obtenons que:

$$\begin{aligned} \bar{\delta} &= \left[\pi [\bar{Y}_1 | D = 1] + (1 - \pi) [\bar{Y}_1 | D = 1] \right] + \\ &\quad \left[\pi [\bar{Y}_0 | D = 0] + (1 - \pi) [\bar{Y}_0 | D = 0] \right] \\ &= [\bar{Y}_1 | D = 1] - [\bar{Y}_0 | D = 0] \end{aligned}$$

- Qui à son tour peut être estimé par l'estimateur correspondant obtenu à partir de l'échantillon :

$$\hat{\delta} = [\hat{Y}_1 | D = 1] - [\hat{Y}_0 | D = 0]$$

- Ainsi, une condition suffisante pour que l'estimateur standard estime systématiquement le véritable ATE est que :

$$[\bar{Y}_1 | D = 1] = [\bar{Y}_1 | D = 0] \quad \text{et} \quad [\bar{Y}_0 | D = 1] = [\bar{Y}_0 | D = 0]$$

- Dans cette situation, l'impact moyen du programme est le même pour le groupes des bénéficiaires et pour le groupe des témoins.
- Afin de satisfaire ces conditions, il suffit que l'affectation au traitement D ne soit pas corrélée aux distributions de résultat potentiel Y_1 et Y_2 .
- La manière la plus facile d'assurer cette non-corrélation est d'affecter les unités de manière aléatoire au traitement.

- Dans la plupart des circonstances, il nous est impossible de savoir comment les membres du groupe témoin auraient réagi s'ils avaient plutôt bénéficié du traitement.
- Ceci forme la base d'un point de vue important dans les biais potentiels de l'estimateur standard (2).

□ Avec un peu d'algèbre, on peut démontrer que :

$$\hat{\delta} = \bar{\delta} + \underbrace{\left([\bar{Y}_0 | D = 1] - [\bar{Y}_0 | D = 0] \right)}_{\text{Différence de base}} + (1 - \pi) \underbrace{\left(\bar{\delta}_{\{D=1\}} - \bar{\delta}_{\{D=0\}} \right)}_{\text{Hétérogénéité de l'effet du traitement}}$$

□ Cette équation spécifie qu'il peut y avoir deux sources de biais dans les estimations des effets causaux faites à partir d'études observationnelles.

1. Le biais de sélection :

les deux groupes sont différents à la base.

2. L'hétérogénéité de l'effet du traitement:

les deux groupes réagissent différemment au traitement

□ La plupart des méthodes d'estimation disponibles ne s'occupent que du biais de sélection, assumant simplement que l'effet de traitement est constant dans la population ou en redéfinissant le paramètre d'intérêt dans la population.

Effet de traitement sur les bénéficiaires

- L'ATE n'est pas toujours le paramètre d'intérêt.
- Dans beaucoup de contextes de politique, c'est l'effet de traitement moyen sur les groupes de bénéficiaires qui nous intéresse :

$$\begin{aligned} \text{TOT} &= E [Y_1(u) - Y_0(u) | D = 1] \\ &= E [Y_1(u) | D = 1] - E [Y_0(u) | D = 1] \end{aligned}$$

Traitement des groupes de bénéficiaires

- L'estimateur standard (2) estime systématiquement TOT à condition que :

$$[\overline{Y}_0 | D = 1] = [\overline{Y}_0 | D = 0]$$

Références

- ❑ Judea Pearl (2000): Causality: Models, Reasoning and Inference, CUP. Chapters 1, 5 and 7.
- ❑ Trygve Haavelmo (1944): “The probability approach in econometrics”, *Econometrica* 12, pp. iii-vi+1-115.
- ❑ Arthur Goldberger (1972): “Structural Equations Methods in the Social Sciences”, *Econometrica* 40, pp. 979-1002.
- ❑ Donald B. Rubin (1974): “Estimating causal effects of treatments in randomized and nonrandomized experiments”, *Journal of Educational Psychology* 66, pp. 688-701.
- ❑ Paul W. Holland (1986): “Statistics and Causal Inference”, *Journal of the American Statistical Association* 81, pp. 945-70, with discussion.