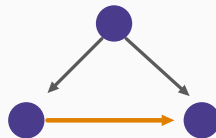


Inférence causale et machine learning

Chapitre 7 : La régression sur discontinuité



L'idée

Les deux variantes

Les vérifications

La mise en œuvre

La portée du résultat

Ce qu'il faut retenir

L'idée

Un seuil administratif comme expérience

La situation

Le traitement est attribué selon qu'une variable continue — la **variable de forçage** — dépasse ou non un **seuil** fixé.

Un score d'éligibilité, un âge, un revenu, un nombre d'habitants.

L'argument, en une phrase

Deux candidats séparés d'un point de score sont **presque identiques** en tout — motivation, aptitude, situation.

Le seuil les sépare pourtant : l'un reçoit le programme, l'autre non. **Au voisinage immédiat du seuil, l'attribution est aussi bonne qu'un tirage au sort.**

Notre programme de formation

Il est réservé aux candidats dont le score d'éligibilité atteint 50.

Position	Taux d'emploi
Juste sous le seuil	40 %
Juste au-dessus	52 %

Les deux variantes

Discontinuité nette

Le seuil détermine **entièrement** le traitement : la probabilité passe de 0 à 1.

L'effet est le saut de la variable de résultat au seuil.

Discontinuité floue

Le seuil **modifie** la probabilité sans la porter à 1 — des éligibles renoncent, des non-éligibles obtiennent une dérogation.

On divise alors le saut du résultat par le saut de la probabilité de traitement. C'est un estimateur de variable instrumentale, que le chapitre 8 formalise.

Le cas flou est le plus fréquent

Peu de dispositifs sont appliqués sans exception. Repérer si la discontinuité est nette ou floue est donc le premier travail — et il se fait en traçant simplement la probabilité de traitement contre la variable de forçage.

Les vérifications

Ce qui rend la stratégie crédible

Trois contrôles indispensables

1. **Continuité des covariables** — les caractéristiques antérieures ne doivent pas sauter au seuil.
2. **Densité de la variable de forçage** — pas d'accumulation anormale d'un côté.
3. **Seuils placebo** — appliquer la méthode à de faux seuils ne doit rien montrer.

La manipulation, menace principale

Si les individus peuvent **contrôler précisément** leur position par rapport au seuil, tout s'effondre : ceux qui parviennent à le franchir diffèrent de ceux qui échouent.

Le symptôme est visible : un **amoncellement** juste au-dessus du seuil, et un creux juste en dessous.

Contrôler approximativement ne suffit pas à invalider

On peut viser un seuil sans l'atteindre exactement — un candidat sait qu'il doit réviser, il ne sait pas s'il obtiendra 49 ou 51.

C'est cette **imprécision résiduelle** qui sauve la méthode : elle recrée du hasard là où l'intention ne suffit pas. La question à poser n'est donc pas « les gens veulent-ils franchir le seuil ? » — évidemment oui — mais « peuvent-ils s'y placer **au point près** ? ».

La mise en œuvre

La fenêtre

On estime **localement**, sur une bande autour du seuil.

Fenêtre	Biais	Variance
Étroite	faible	élevée
Large	élevé	faible

C'est le compromis biais-variance du chapitre 2 d'*Apprentissage statistique*, sous une forme nouvelle.

La forme fonctionnelle

On ajuste séparément de chaque côté du seuil, généralement par un polynôme **de degré un ou deux**.

Les degrés élevés produisent des artefacts au bord et sont aujourd'hui déconseillés.

Définition

Le résultat doit être présenté pour **plusieurs fenêtres** et plusieurs degrés. Un effet qui disparaît quand on change légèrement la fenêtre n'est pas un effet.

La portée du résultat

Un effet local, et il faut le dire

Effet local au seuil

La régression sur discontinuité identifie l'effet **pour les individus situés au voisinage du seuil**, et pour eux seulement.

Ce que cela implique concrètement

Notre effet de douze points concerne les candidats dont le score avoisine 50.

Il ne dit **rien** de ce que le programme ferait à quelqu'un dont le score est de 20 — dont les besoins et la capacité à en tirer parti sont sans doute très différents.

Répondre à « faut-il abaisser le seuil ? » exige donc de **l'extrapoler**, et cette extrapolation n'est pas garantie par la méthode.

L'arbitrage entre crédibilité et portée

	Crédibilité	Portée
Discontinuité	très élevée	étroite
Appariement (ch. 5)	faible	large

Ce qu'il faut retenir

Cinq points

1. Un **seuil** administratif sépare des individus quasi identiques : au voisinage, l'attribution est quasi aléatoire.
2. Sur notre programme : $52 - 40 = \mathbf{12}$ points d'effet local.
3. **Nette** ou **floue** selon que le seuil détermine ou modifie seulement le traitement.
4. Vérifier la **continuité des covariables**, la **densité**, et des **seuils placebo**; la manipulation est la menace principale.
5. L'effet identifié est **local au seuil** — grande crédibilité, portée étroite.

La suite

La discontinuité floue a introduit une division par un saut de probabilité. Le chapitre 8 généralise cette idée : les variables instrumentales.