SFC-CLADAG meeting
Caserta 11-13 june 2008

Causal inference through principal stratification: a special type of latent class modelling

Leonardo Grilli

grilli@ds.unifi.it
www.ds.unifi.it/grilli

Department of Statistics

University of Florence

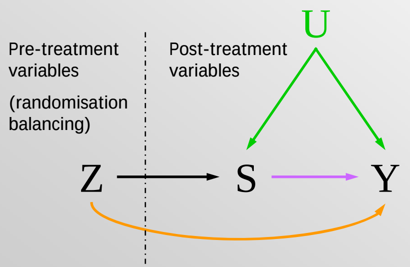
Outline

- Causal inference with intermediate variables: principal stratification
- An application: comparing the effectiveness of two degree programmes
- Connection between principal stratification and latent class modelling

Causal inference with intermediate variables:

principal stratification

Causal inference in the presence of a relevant intermediate variable



Z = treatment

S = intermediate

Y = response

U = unobs. variables

1) **Non-compliance**: Z assignment, S actual treatment
2) **Direct/indirect effects**: disentangle total effect of Z on Y into direct effect and indirect effect through S
3) **Missing response**: Y observed or not depending on S
4) **Truncation due to death**: Y exists or does not exist depending on S

Causal inference in the presence of a relevant intermediate variable

- Under randomisation (more generally, under unconfoundedness) all pre-treatment variables are balanced among treatment groups
- But it is not so for the post-treatment variables
- This fact may lead to biases if the post-treatment variable is *relevant*, e.g. if one wishes to
  - condition on a post-treatment variable, or
  - estimate a causal effect for a post-treatment variable

→ Possible solution: define **potential outcomes** for all the post-treatment variables

Potential outcomes

Causal inference:

many approaches, *much controversy*

- here I refer to Rubin's **potential outcomes** approach

- For each *relevant* post-treatment variable there is one potential version for each level of the treatment
- Every statistical unit is assigned to one and only one level of the treatment, so only one of the potential versions is observable
- If Z binary indicator of treatment then
  - S(1), S(0) $\Rightarrow$ S=S(Z) observed version
  - Y(1), Y(0) $\Rightarrow$ Y=Y(Z) observed version

Principal strata

Simplest case: both Z and S dichotomous $\rightarrow$ 4 strata

Z=1 drug

S=1 get well

Z	sicker	opposite	responsive	healthier
1	**0**	**0**	**1**	**1**
0	**0**	**1**	**0**	**1**

Principal strata are defined by the values of the two potential versions of the intermediate variable S (counterfactual) → they are not influenced by the value taken by Z (like pre-treatment covariates)

Observed values of Z and S do not identify the stratum: if Z=1 and S=1 the unit can belong to two strata: 10 (responsive) or 11 (healthier)

Principal strata are latent classes (→ ***latent class models***)

Principal causal effects

Causal effect of Z on Y for a single unit: $Y_i(1)$ vs. $Y_i(0)$

Principal causal effect of Z on Y:

$f(Y(1))$ vs. $f(Y(0))$ for the units of a principal stratum

Causal effects across principal strata are nonsense

Conditioning on the *observed* value of the intermediate variable S implies conditioning on different principal strata depending on the value of Z

Refs

Frankgakis C.E. & Rubin D.B. (2002) Principal stratification in causal inference, *Biometrics*, 58: 21-29.

Barnard J., Frangakis C.E., Hill J.L. & Rubin D.B. (2003) Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City, *JASA*, 98: 299-323.

Truncation due to death

Z=treatment, S=survival, Y=quality of life

→ Y defined only for S=1 (no quality of life for dead persons!)

BUT: non-sense to compare Y under Z=0 and Z=1 among the survivors (i.e. condition on S=1):

Z=0 and S=1 ⇔ unit ∈ strata 11 or 01

Z=1 and S=1 ⇔ unit ∈ strata 11 or 10

outcome not defined under one value of Z → causal effect undefined

The only conceivable casual effect of Z on Y is the principal effect in the stratum 11, namely {S(0)=1, S(1)=1}

Zhang J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by 'death', JEBS *28*, pp. 353-368.

An application:

comparing the effectiveness of two degree programmes

Overview

Joint work with Fabrizia Mealli, Univ. of Florence

AIM: assessing the relative effectiveness of two degree programmes with respect to employment

- 1992 cohort of freshmen of the University of Florence
- Enrolment in two degree programmes: Economics (Z=1) and Political Science (Z=0)
- Graduation: binary indicator for graduation (S=1)
- Employment: binary indicator for having a permanent job two years after degree (Y=1)

The indicator for graduation (S=1) is a relevant intermediate variable

Data

A. **Administrative database** of the 1992 cohort of freshmen enrolled in *Economics* (1068 students) and *Political Science* (873 students)

B1-B3. Three **census surveys** on the occupational status of the graduates of the University of Florence of years 1998 to 2000

datasets A and B1-B3 are merged

Available covariates: Female, Residence in Florence, Gymnasium (Lyceum), High grade, Late enrolment

covariates are important since the treatment is not randomised!

Variables /1

Treatment variable Z:

$$Z = \begin{cases} 1 & \text{if enrolled in Economics} \\ 0 & \text{if enrolled in Political Science} \end{cases}$$

- Z is called "treatment" just to conform to the literature on causal inference
- No active vs. placebo → values of Z on an equal footing
- No randomisation → possible confounders (so covariates are important for unconfoundedness)

Variables /2

Intermediate variable S:

$$S = S(z) = \begin{cases} 1 & \text{if graduated when } z \\ 0 & \text{if not graduated when } z \end{cases}$$

S is the observed version of the potential variables S(0), S(1)

Outcome variable Y:

$$Y = Y(z) = \begin{cases} 1 & \text{if employed (after graduation) when } z \\ 0 & \text{if not employed (after graduation) when } z \end{cases}$$

Y is the observed version of the potential outcomes Y(0), Y(1)

For our purposes Y is defined only when S=1

Principal strata

In our case both Z and S are binary → 4 strata

Z	L=GG	L=GN	L=NG	L=NN
1 (Economics)	G	G	N	N
0 (Political Sc)	G	N	G	N

G=Graduated

N=Not graduated

Principal strata are defined by values of the two potential versions of the intermediate var. S (counterfactual): e.g. **GN** are the students who become **G**raduate if enrolled in Economics and **N**ot graduate if enrolled in Political Sc.

Observed group $O(Z, S^{obs})$	Z_i	S_i^{obs}	Y_i^{obs}	Latent group L_i (principal stratum)
$O(1,1)$	1	1	in {0,1}	GG or GN
$O(1,0)$	1	0	not defined	NG or NN
$O(0,1)$	0	1	in {0,1}	GG or NG
$O(0,0)$	0	0	not defined	GN or NN

mixtures

Relevant parameters

Probabilities of the principal strata: $\pi_{GG}, \pi_{GN}, \pi_{NG}, \pi_{NN}$

e.g. probability to be a student who become **G**raduate if enrolled in Economics and **N**ot graduate if enrolled in Political Science

Probabilities of employment: $\gamma_{1,GG}, \gamma_{0,GG}, \gamma_{1,GN}, \gamma_{0,NG}$

e.g. probability to be employed for a student who (i) become **G**raduate if enrolled in Economics and **N**ot graduate if enrolled in Political Science and (ii) actually enrolled in Economics

Causal effect of degree prog. on employment in the GG group: $\gamma_{1,GG} - \gamma_{0,GG}$

Statistical techniques

Principal stratification is the conceptual framework for the application of various statistical techniques:

- **Non parametric methods** (**bounds**)
  Grilli L. & Mealli F. (2008) JEBS
- **Model-based methods** (**latent class models**)
  - ML
  - Bayesian

In this talk I show the ML results:

Grilli L. & Mealli F. (2007) University Studies and Employment. An Application of the Principal Strata Approach to Causal Analysis. In: Effectiveness of University Education in Italy (L. Fabbris ed.). Springer Verlag.

Likelihood

$$L\left(\boldsymbol{\theta} \mid \mathbf{Z}, \mathbf{S}^{obs}, \mathbf{Y}^{obs}, \mathbf{X}\right) =$$

Enrolled in Economics:

$$\prod_{i \in O(1,1)} \left\{ \pi_{GG:i} \left[\left(\gamma_{1,GG:i}\right)^{Y_i^{obs}} \left(1-\gamma_{1,GG:i}\right)^{1-Y_i^{obs}} \right] + \pi_{GN:i} \left[\left(\gamma_{1,GN:i}\right)^{Y_i^{obs}} \left(1-\gamma_{1,GN:i}\right)^{1-Y_i^{obs}} \right] \right\}$$

$$\times \prod_{i \in O(1,0)} \left\{ \pi_{NG:i} + \pi_{NN:i} \right\}$$

Enrolled in Political Sc.:

$$\times \prod_{i \in O(0,1)} \left\{ \pi_{GG:i} \left[\left(\gamma_{0,GG:i}\right)^{Y_i^{obs}} \left(1-\gamma_{0,GG:i}\right)^{1-Y_i^{obs}} \right] + \pi_{NG:i} \left[\left(\gamma_{0,NG:i}\right)^{Y_i^{obs}} \left(1-\gamma_{0,NG:i}\right)^{1-Y_i^{obs}} \right] \right\}$$

$$\times \prod_{i \in O(0,0)} \left\{ \pi_{GN:i} + \pi_{NN:i} \right\}$$

Various models can be built by specifying submodels for the π's and the γ's

Model specification

Probabilities of the principal strata: π_{GG}, π_{GN}, π_{NG}, π_{NN}

Principal strata submodel: multinomial logit

Probabilities of employment: $\gamma_{1,GG}$, $\gamma_{0,GG}$, $\gamma_{1,GN}$, $\gamma_{0,NG}$

Outcome submodel: 4 separate logit models

Principal strata are latent classes

$\Rightarrow$ the model is a *latent class model* with restrictions:

- each individual (given Z and S) can belong to only 2 of the 4 classes
- the outcome is not defined for some classes (depending on Z)

Principal strata submodel

$$\pi_{GG:i} = \frac{\exp(\eta^{\pi}_{GG:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

$$\pi_{GN:i} = \frac{\exp(\eta^{\pi}_{GN:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

$$\pi_{NG:i} = \frac{\exp(\eta^{\pi}_{NG:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

$$\pi_{NN:i} = \frac{1}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

Multinomial logit specification

$$\eta^{\pi}_{GG:i} = \alpha^{\pi}_{GG} + \boldsymbol{\beta}^{\pi}_{GG}{}'\mathbf{x}_i$$
$$\eta^{\pi}_{GN:i} = \alpha^{\pi}_{GN} + \boldsymbol{\beta}^{\pi}_{GN}{}'\mathbf{x}_i$$
$$\eta^{\pi}_{NG:i} = \alpha^{\pi}_{NG} + \boldsymbol{\beta}^{\pi}_{NG}{}'\mathbf{x}_i$$

With 5 covariates there are $3+3\times5=18$ parameters

Outcome submodel

$$\gamma_{1,GG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{1,GG:i})}$$

$$\gamma_{0,GG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{0,GG:i})}$$

$$\gamma_{1,GN:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{1,GN:i})}$$

$$\gamma_{0,NG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{0,NG:i})}$$

Separate logit specifications

$$\eta^{\gamma}_{1,GG:i} = \alpha^{\gamma}_{1,GG} + \boldsymbol{\beta}^{\gamma\prime}\mathbf{x}_i$$

$$\eta^{\gamma}_{0,GG:i} = \alpha^{\gamma}_{0,GG} + \boldsymbol{\beta}^{\gamma\prime}\mathbf{x}_i$$

$$\eta^{\gamma}_{1,GN:i} = \alpha^{\gamma}_{1,GN} + \boldsymbol{\beta}^{\gamma\prime}\mathbf{x}_i$$

$$\eta^{\gamma}_{0,NG:i} = \alpha^{\gamma}_{0,NG} + \boldsymbol{\beta}^{\gamma\prime}\mathbf{x}_i\,.$$

With 5 covariates there are 4+5=9 parameters

ML inference

- ✓ Maximization algorithm: quasi-Newton with a BFGS update of the Cholesky factor of the approximate Hessian
- ✓ Software: SAS proc NLMIXED

- Principal strata submodel ⇒ 18 parameters
- Outcome submodel ⇒ 9 parameters

Overall 27 parameters

Some parameters of the Principal strata submodel (a multinomial logit) have highly negative estimates and huge standard errors
⇒ for certain values of the covariates some principal strata are empty so some constraints are needed (the final model has 8 constraints)

Estimated probabilities (%) for some covariate patterns

Parameter	00000	00100	00110	00101	01100	10100	11100	11111
$\pi_{GG:i}$	1.1	8.0	9.1	10.9	20.3	24.9	52.5	62.2
$\pi_{GN:i}$	6.3	6.0	3.3	0.0	14.0	0.0	0.0	0.0
$\pi_{NG:i}$	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\pi_{NN:i}$	89.0	86.0	87.6	89.1	65.7	75.1	47.5	37.8
$\gamma_{1,GG:i}$	77.9	58.2	59.9	60.7	57.3	48.0	47.1	51.5
$\gamma_{0,GG:i}$	64.5	41.7	43.4	44.2	40.8	32.2	31.4	35.3
$\gamma_{1,GN:i}$	61.9	39.0	40.7	41.5	38.1	29.8	29.0	32.8
$\gamma_{0,NG:i}$	20.3	9.1	9.7	10.0	8.9	6.3	6.1	7.1
Causal effect $\gamma_{1,GG:i} - \gamma_{0,GG:i}$	13.5	16.5	16.5	16.4	16.5	15.8	15.7	16.2

The pattern $(x_1, x_2, x_3, x_4, x_5)$ stands for

Gymnasium $= x_1$, *High grade* $= x_2$, *Regular enrolment* $= x_3$, *Female* $= x_4$, *Florence* $= x_5$

Connection between principal stratification and latent class modelling

Principal strata and latent class modelling

- A parametric model derived within the principal strata framework is a special instance of latent class model
- The connection was recognized by Bengt Muthén in the case of non-compliance (CACE: Complier Average Casual Effect)
  - Muthén B. (2002) Beyond SEM: general latent variable modeling, *Behaviormetrika*.
  - Mplus user's guide (**`www.statmodel.com`**) with a re-analysis of Little & Yau (1998) data
- In the software Mplus the class membership restrictions are handled by training data, i.e. an auxiliary dataset that reports for each sample unit which classes are admissible and which classes are not

Principal strata and latent class modelling

- The connection between principal strata and latent class modelling is exploited also by Skrondal & Rabe-Hesketh (2004) in their book Generalized Latent Variable Modeling
  - They show how a CACE model can be written as a latent class model that fits the GLLAMM framework
  - They re-analyse Little & Yau (1998) data using the Stata **`gllamm`** command

While the connection is recognized in the non-compliance case (CACE), there has been no discussion of the connection in the more general principal stratification framework. Also the implications of the connection have not been investigated.

Peculiarities of latent class models derived from principal strata

1. The number of classes and their meaning is determined a priori, as each class corresponds to a principal stratum
   - avoid the tricky problem of a data-driven choice of the number of latent classes
   - avoid the somewhat arbitrary exercise of attaching labels to the classes
2. An individual can only belong to a subset of latent classes, i.e. given the data the probabilities of belonging to certain classes are zero by assumption
   - estimation is simpler with respect to a standard LC model with the same number of classes, since some components of the mixtures are ruled out by assumption

Peculiarities of latent class models derived from principal strata

Truncation by death adds another peculiarity:

3. Latent class membership determines whether the outcome is defined or not (and its probability in case it is defined)
   - this feature is specific to truncation by death in the principal strata framework and does not apply to standard LC models, where it is not conceivable to let the outcome be defined or not depending on the class

Peculiarities of latent class models derived from principal strata

- As for model specification, principal stratification gives solid arguments to put restrictions on the latent classes based on
  - substantive assumptions: e.g. in experiments with non-compliance the latent class of *defiers* can be assumed to be empty based on considerations on the behaviour of the individuals (monotonicity)
  - design: e.g. the latent class of *always takers* is empty if the design prevents people assigned to control from taking the active treatment
- Last but not least, a LC model with a structure derived within the principal strata framework guarantees that the model is consistent with the principles of counterfactual causal inference and thus the parameters refer to well-defined causal quantities

- Questions and further material
  - Email: **grilli@ds.unifi.it**
  - Web: **www.ds.unifi.it/grilli**

Principal strata submodel results

	Initial model		*Final model*	
Number of parameters	27		21	
Deviance (-2*logL*)	2231.8		2231.8	
Principal strata submodel (π's)				
α^{π}_{GG}	-4.403	(0.449)	-4.402	(0.448)
α^{π}_{GN}	-2.644	(0.749)	-2.647	(0.752)
α^{π}_{NG}	-3.206	(0.836)	-3.207	(0.835)
$\beta^{\pi}_{GG,gymnasium}$	1.275	(0.157)	1.275	(0.157)
$\beta^{\pi}_{GN,gymnasium}$	-5.757	(n.a.)	$-\infty$	
$\beta^{\pi}_{NG,gymnasium}$	-15.041	(n.a.)	$-\infty$	
$\beta^{\pi}_{GG,high\_grade}$	1.204	(0.146)	1.205	(0.146)
$\beta^{\pi}_{GN,high\_grade}$	1.113	(0.653)	1.113	(0.652)
$\beta^{\pi}_{NG,high\_grade}$	-8.092	(114.022)	$-\infty$	
$\beta^{\pi}_{GG,regular\_enrolment}$	2.024	(0.425)	2.023	(0.425)
$\beta^{\pi}_{GN,regular\_enrolment}$	-0.012	(0.788)	-0.009	(0.792)
$\beta^{\pi}_{NG,regular\_enrolment}$	-8.140	(64.473)	$-\infty$	
$\beta^{\pi}_{GG,female}$	0.117	(0.137)	0.117	(0.137)
$\beta^{\pi}_{GN,female}$	-0.617	(0.753)	-0.622	(0.755)
$\beta^{\pi}_{NG,female}$	0.988	(1.112)	0.991	(1.111)
$\beta^{\pi}_{GG,Florence}$	0.280	(0.144)	0.280	(0.144)
$\beta^{\pi}_{GN,Florence}$	-13.499	(559.599)	$-\infty$	
$\beta^{\pi}_{NG,Florence}$	-10.353	(533.855)	$-\infty$	

Principal strata submodel results

- the size of GG stratum varies a lot with the covariates, from a minimum of 1.1% (students with weak background) to a maximum of 62.2%
- for most covariate patterns the GN and NG strata (i.e. students able to graduate in only one degree prog.) are very small (but for students with weak background they are larger then the GG stratum)
- the higher graduation rate of Economics is originated by the students with a weak background ⇒ orientation policies should be designed especially for this kind of students

Outcome submodel results

	Initial model		Final model	
Number of parameters		27		21
Deviance (-2*logL*)		2231.8		2231.8
Outcome submodel (γ's)				
$\alpha^{\gamma}_{1,GG}$	1.257	(1.240)	1.262	(1.241)
$\alpha^{\gamma}_{0,NG}$	-1.357	(1.561)	-1.365	(1.568)
$\alpha^{\gamma}_{0,GG}$	0.593	(1.185)	0.596	(1.185)
$\alpha^{\gamma}_{1,GN}$	0.498	(1.057)	0.484	(1.058)
$\beta^{\gamma}_{gymnasium}$	-0.405	(0.374)	-0.410	(0.374)
$\beta^{\gamma}_{high\_grade}$	-0.035	(0.262)	-0.036	(0.263)
$\beta^{\gamma}_{regular\_enrolment}$	-0.933	(0.979)	-0.932	(0.979)
β^{γ}_{female}	0.072	(0.272)	0.070	(0.272)
$\beta^{\gamma}_{Florence}$	0.106	(0.333)	0.104	(0.333)
Causal effect $\alpha^{\gamma}_{1,GG} - \alpha^{\gamma}_{0,GG}$	0.664	(0.301)	0.666	(0.301)

Outcome submodel results

- the level of the probability of being employed varies a lot with the covariates
- in the GG stratum the causal effect on employment (modelled as constant across the covariate patterns) is about 15% (significant at 5%)
- students with a weak background have little chances of being GG, so for them the above causal effect has little relevance