

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

Leonardo Grilli

Department of Statistics ‘Giuseppe Parenti’, University of Florence  
grilli@ds.unifi.it

November 15, 2010

## Abstract

Principal stratification is an increasingly adopted framework for drawing counterfactual causal inferences in complex situations. After outlining the framework, with special emphasis on the case of truncation by death, I describe an application of the methodology where the analysis is based on a parametric model with latent classes. Then, I discuss the special features of latent class models derived within the principal strata framework. I argue that the concept of principal stratification gives latent class models a solid theoretical basis and helps to solve some specification and fitting issues.

## 1 Outline

Principal stratification is a conceptual framework developed in the setting of counterfactual causal inference to deal with situations where the causal path from the treatment to the outcome includes an intermediate variable that cannot be ignored (3). Examples are *non-compliance* (1; 2; 10), where the intermediate variable is the compliance status, estimation of *direct effects* (14; 16), where the intermediate variable is a variable whose effect one wishes to control for, *surrogacy in clinical trials* (7; 8), where the intermediate variable is a surrogate endpoint, and *truncation by death* (18; 4; 5; 19), where the intermediate variable determines the existence of the outcome.

Basically, the problem with intermediate variables is that they are measured after treatment and thus they are not balanced among the treatment arms. Therefore, the conventional estimators of the causal effect of an intermediate variable are generally biased; moreover, conditioning on an intermediate variable may bias the estimators of other causal effects of interest.

The application described in the next section focuses on truncation by death, a case taking its name from the studies on the quality of life, where the outcome of interest does not exist for patients who died. The simplest approach is to carry out the analysis on the patients who survived, but this is likely to yield biased results since survival is a post-treatment variable and conditioning on it destroys the randomized structure of the experiment. Zhang and Rubin (18) noted that the same issue may arise in experiments for comparing educational programmes. In fact, they applied principal stratification to the hypothetical case of a randomized experiment concerning two high school educational programmes, where the

intermediate variable is graduation and the outcome is the score on a final test. This is an instance of truncation by death since the outcome of interest exists only for students who graduated.

Later Grilli and Mealli (4; 5) used principal strata to tackle a case of truncation by death in the evaluation of the effectiveness of two degree programmes with respect to job opportunities, where the treatment is the degree programme (Economics vs Political Science), the intermediate variable is the graduation status (graduated within 9 years) and the outcome is the employment status (having a permanent job). This is another instance of truncation by death: since the aim is to assess the relative effectiveness of graduation in different degree programmes, the employment status is not defined for students who did not graduate.

A further application of principal stratification to deal with truncation by death is given by Zhang, Rubin and Mealli (19) in the context of the effectiveness of job-training programs: indeed, estimating the effects of training programs on wages is complicated by the fact that, even in a randomized experiment, wages are truncated by nonemployment, that is, they are only observed and well-defined for individuals who are employed.

The paper proceeds with a section illustrating principal stratification through an application to the effectiveness of degree programmes and a section discussing the latent class perspective of principal stratification.

## 2 Review of principal stratification

The principal stratification framework requires a treatment with a finite number of levels and two post-treatment variables, namely an intermediate variable and an outcome. The nature of the intermediate variable determines the type of strata: a discrete intermediate variable implies discrete strata, while a continuous intermediate variable implies continuous strata. It will be clear that only discrete strata can be seen as latent classes. The simplest case of discrete principal strata arises when both the treatment and the intermediate variable are binary, implying four principal strata.

The principal stratification framework will be illustrated through the application of Grilli and Mealli (4; 5), who analyzed 1941 freshmen of the University of Florence: 1068 enrolled in Economics and 873 in Political Science.

The *treatment*  $Z_i$  takes the value 1 if student  $i$  enrolled in Economics and 0 if enrolled in Political Science. Under the standard Stable Unit Treatment Value Assumption (3) (SUTVA), the post-treatment variables are defined as follows. The *intermediate variable*  $S_i(z_i)$  is 1 or 0 if student  $i$  graduated or did not graduate within 9 years when enrolled in degree programme  $z_i$ . The *outcome*  $Y_i(z_i)$  is 1 or 0 if student  $i$  had or did not have a permanent job at the time of the interview (i.e. from one to two years after the degree) when enrolled in programme  $z_i$  and graduated.

Since for each individual the treatment assumes a single value, for every post-treatment variable only one of the two potential versions can be observed:  $S_i^{obs} = S_i(Z_i)$  and  $Y_i^{obs} = Y_i(Z_i)$ . Since both the treatment and the intermediate variable are binary, there are four *principal strata*:

- $GG$  (Graduated, Graduated) if  $S_i(1)=1$  and  $S_i(0)=1$ ;
- $GN$  (Graduated, Not graduated) if  $S_i(1)=1$  and  $S_i(0)=0$ ;
- $NG$  (Not graduated, Graduated) if  $S_i(1)=0$  and  $S_i(0)=1$ ;
- $NN$  (Not graduated, Not graduated) if  $S_i(1)=0$  and  $S_i(0)=0$ .

The principal stratum of individual  $i$  cannot be observed since either  $Z_i = 0$  or  $Z_i = 1$ . The principal stratum is thus a latent class, denoted with a latent variable  $C_i$  taking values in the set  $\{GG, GN, NG, NN\}$ . The probability that an individual belongs to a given principal stratum can be estimated. A crucial feature is that, given the values of the treatment  $Z_i$  and the intermediate variable  $S_i^{obs}$ , some principal strata are ruled out: for example, a student who enrolled in Economics ( $Z_i = 1$ ) and then graduated ( $S_i^{obs} = 1$ ) can only belong to the strata  $GG$  and  $GN$ , so the strata  $NG$  and  $NN$  are inadmissible and their probability is null.

The key feature of the principal strata is that they are defined by the couple of potential values of the intermediate variable, so they are not affected by the treatment and thus can be seen as categories of an unobserved pre-treatment covariate.

The terms entering the causal effect of interest  $Y_i(1) - Y_i(0)$  are both defined only in the  $GG$  stratum, i.e. students who would be able to graduate in both programmes. The estimand of main interest is thus the *Average Causal Effect (ACE) on employment in the GG stratum*, i.e. the difference between the probabilities of being employed for Economics and Political Science in the subset of students that would be able to graduate in any of the two degree programmes.

Grilli and Mealli (4; 5) included also some covariates  $\mathbf{x}_i$ . In general, covariates are important when the treatment is not randomized, since the *unconfoundedness assumption* required for the causal interpretation of the effect of the treatment is more reasonable if stated conditional on good covariates. Formally, the treatment is conditionally unconfounded when  $Z_i \perp \{S_i(0), S_i(1), Y_i(0), Y_i(1)\} | \mathbf{x}_i$ .

Under the assumptions of SUTVA and conditional unconfoundedness, the data generating process can be defined in terms of two sets of probabilities: the probabilities of the principal strata and the probabilities of the outcome conditional on the principal stratum.

The probabilities of the principal strata are

$$\{\pi_{GG:i}, \pi_{GN:i}, \pi_{NG:i}, \pi_{NN:i}\},$$

e.g.  $\pi_{GG:i} = \Pr(C_i = GG | \mathbf{x}_i)$ .

The probabilities of the outcome conditional on the principal stratum are

$$\{\gamma_{1,GG:i}, \gamma_{0,GG:i}, \gamma_{1,GN:i}, \gamma_{0,GN:i}\},$$

where the number 0 or 1 in the subscript is the value of  $Z_i$ . For example,  $\gamma_{0,GG:i} = \Pr(Y_i(0) = 1 | C_i = GG, \mathbf{x}_i)$ . Here the  $\gamma$ 's for other combinations of programme and principal stratum, such as  $Z_i = 1$  and  $C_i = NG$ , are not defined.

As in the majority of the applications with principal strata, the treatment and the intermediate variable are both binary, leading to four principal strata. However, while in many settings it is sensible to assume that certain strata are empty (e.g. the assumption of no defiers in an experiment with non-compliance), in the present context such assumptions are not plausible in the light of the symmetry of the two treatments, so all the strata are allowed to exist and thus every observed group is generated by a mixture of two distributions.

The principal stratification framework can be exploited to carry out a non-parametric analysis based on large-sample bounds (5) or to build a parametric model to be fitted with Bayesian or likelihood methods (4). In the example, the likelihood is a product over four

observable groups defined by  $Z_i$  and  $S_i^{obs}$ :

$$\prod_{i: Z_i=1, S_i^{obs}=0} \{\pi_{NG:i} + \pi_{NN:i}\} \times \prod_{i: Z_i=1, S_i^{obs}=1} \{\pi_{GG:i}B_{1,GG:i} + \pi_{GN:i}B_{1,GN:i}\} \times \\ \prod_{i: Z_i=0, S_i^{obs}=0} \{\pi_{GN:i} + \pi_{NN:i}\} \times \prod_{i: Z_i=0, S_i^{obs}=1} \{\pi_{GG:i}B_{0,GG:i} + \pi_{NG:i}B_{0,NG:i}\}$$

where the  $B$ 's are the Bernoulli likelihoods for the  $\gamma$ 's, for example

$$B_{1,GG:i} = (\gamma_{1,GG:i})^{Y_i^{obs}} (1 - \gamma_{1,GG:i})^{1-Y_i^{obs}}.$$

The parametric model devised by Grilli and Mealli (4) is made of two components: a multinomial logit model for the probabilities of the principal strata conditional on the covariates (the  $\pi$ 's) and a set of logit models for the probabilities of the outcome conditional on both the covariates and the principal stratum (the  $\gamma$ 's). The model is thus a latent class model, but the principal stratification framework entails some peculiarities that make the analysis different from traditional latent class modelling.

### 3 Principal stratification and latent class modelling

In the previous section it has been shown that in the case of discrete principal strata the corresponding statistical model is a latent class (LC) model. Note that even if almost all applications assume discrete strata, the principal strata can also be continuous: for example, Jin and Rubin (6) tackled partial compliance by defining the strata as couples of proportion of compliance to drug and proportion of compliance to placebo.

The connection between principal stratification and LC modelling has been recognized in the case of *non-compliance*, with reference to the simple instance of a binary treatment and a binary compliance status (all-or-none compliance). In the notation of the previous section, the intermediate variable  $S_i(z_i)$  is the compliance status under treatment  $z_i$ . The target quantity, called Complier Average Causal Effect (CACE), is the average difference  $Y_i(1) - Y_i(0)$  for individuals in the principal stratum of compliers, namely the individuals that comply with the treatment regardless of the assigned treatment (1).

Bengt Muthén described CACE modelling in terms of LC modelling in (11) and then implemented the idea in the *Mplus* software, whose user's manual (12) reports a re-analysis of Little and Yau's data (9). In *Mplus* the class membership restrictions are handled by the so-called *training data*, i.e. an auxiliary dataset declaring, for each sample unit, which classes are admissible and which classes are not. The possibility to specify a CACE model as an LC model with restrictions is also noted by Vermunt and Magidson in the manual of the software Latent GOLD (17), where the class membership restrictions are inserted via the *Known Class* option.

The latent class perspective in CACE modelling was exploited also by Skrondal and Rabe-Hesketh in their book on Generalized Linear Latent Mixed Models (15), where they showed how a CACE model can be written as an LC model that fits the GLLAMM framework. Moreover, they re-analyzed Little and Yau's data using the Stata `gllamm` command (13).

The mentioned treatments of CACE via LC models are aimed at showing that causal inference can be carried out within a general statistical modelling framework based on latent variables. However, the implications of the connection have not been investigated.

Moreover, there seems to be no discussion of the connection in the more general principal stratification framework, thus including topics such as direct effects and truncation by death.

Let us use the notation introduced in the previous section and let us denote the latent class corresponding to the principal stratum with  $C_i = c$  for  $c \in \mathcal{C}$ . An LC model derived within a framework with discrete principal strata differs from a general LC model in several respects: (a) the number of classes (i.e. the cardinality of  $\mathcal{C}$ ) and their meaning is determined a priori, as each class corresponds to a principal stratum; (b) 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:  $\exists c \in \mathcal{C}$  such that  $\Pr(C_i = c | Z_i, S_i^{obs}, \mathbf{x}_i) = 0$ . Truncation by death adds another peculiarity, namely: (c) latent class membership determines whether the outcome is defined or not (and its probability in case it is defined):  $\exists c \in \mathcal{C}$  such that  $Y_i(z_i)$  is not defined.

Feature *a* allows to avoid the tricky problem of a data-driven choice of the number of latent classes and the somewhat arbitrary exercise of attaching labels to the classes. Feature *b* makes estimation 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. Feature *c* 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.

As for model specification, principal stratification gives solid arguments to put restrictions on the latent classes based on substantive assumptions or on the design: for example, in experiments with non-compliance (1; 2) the latent class of *defiers* can be assumed to be empty based on considerations on the behaviour of the individuals, while 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 latent class 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.

## References

- [1] Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996) Identification of Causal Effects Using Instrumental Variables, *J. Am. Stat. Assoc.* 91, pp. 444–472.
- [2] Barnard, J., Frangakis, C. E., Hill, J. L. and Rubin, D. B. (2003) Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City, *J. Am. Stat. Assoc.* 98, pp. 299–323.
- [3] Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference, *Biometrics* 58, pp. 21–29.
- [4] Grilli, L. and Mealli, F. (2007) University Studies and Employment. An Application of the Principal Strata Approach to Causal Analysis, in *Effectiveness of University Education in Italy*, ed. by L. Fabbri, pp. 219–232. Heidelberg: Physica-Verlag.
- [5] Grilli, L. and Mealli, F. (2008) Nonparametric Bounds on the Causal Effect of University Studies on Job Opportunities Using Principal Stratification. *J. Educ. Behav. Stat.* 33, pp. 111–130.

- [6] Jin, H. and Rubin, D. B. (2008) Principal Stratification for Causal Inference With Extended Partial Compliance, *J. Am. Stat. Assoc.* 103, pp. 101–111.
- [7] Joffe, M. M. and Greene, T. (2009). Related causal frameworks for surrogate outcomes, *Biometrics* 65, pp. 530-538.
- [8] Li, Y., Taylor, J. M. G. and Elliott, M. R. (2010). A Bayesian Approach to Surrogacy Assessment Using Principal Stratification in Clinical Trials, *Biometrics* 66, pp. 523-531.
- [9] Little, R. J. and Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubins causal model, *Psychol. Methods* 3, pp. 147–159.
- [10] Mattei, A. and Mealli, F. (2007). Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-Examination, *Biometrics* 63, 437-446.
- [11] Muthén, B. O. (2002). Beyond SEM: general latent variable modeling, *Behaviormetrika* 29, pp. 81–117.
- [12] Muthén, L. K. and Muthén, B. O. (2007). *Mplus User's Guide. Fifth Edition.*. Los Angeles, CA: Muthén & Muthén.
- [13] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004). GLLAMM Manual, *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.*
- [14] Rubin, D. B. (2004). Direct and Indirect Causal Effects via Potential Outcomes, *Scand. J. Stat.* 31, pp. 161–170.
- [15] Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models.* Boca Raton, FL: Chapman & Hall/ CRC Press.
- [16] VanderWeele, T. (2008). Simple relations between principal stratification and direct and indirect effects, *Stat. Probabil. Lett.* 78, pp. 2957-2962.
- [17] Vermunt, J. K. and Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced.* Belmont, MA: Statistical Innovations Inc.
- [18] Zhang, J. L. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by 'death', *J. Educ. Behav. Stat.* 28, pp. 353–368.
- [19] Zhang, J. L., Rubin, D. B. and Mealli F. (2009). Likelihood-Based Analysis of Causal Effects of Job-Training Programs Using Principal Stratification, *J. Am. Stat. Assoc.* 104, pp. 166–176.