IWSM 2010 – Glasgow

Tuesday 6th July 2010 - Poster Session 1

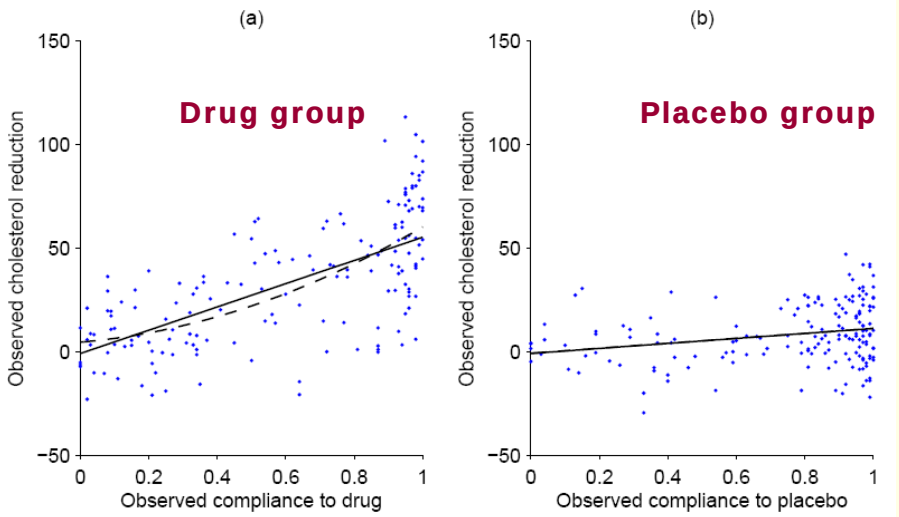
Likelihood inference for a semi-parametric causal model addressing *partial compliance* by continuous *principal strata*

Francesco Bartolucci
University of Perugia
bart@stat.unipg.it

Leonardo Grilli (presenting author)
University of Florence
grilli@ds.unifi.it

1. Efron-Feldman data

- Subset of data from LRC-CPPT, a *placebo-controlled double-blinded randomized clinical trial* designed to study the effectiveness of *cholestyramine* for lowering cholesterol levels
- data on 335 men: 164 → active pills of the drug 171 → placebo pills
  - binary indicator for treatment assignment
  - proportion of compliance (based on pills taken)
  - continuous outcome variable: average decrease in the cholesterol level during the study (average 7.3 years)



observed compliance to placebo
larger than
observed compliance to drug
(adverse side-effects of the drug)

EF imputed the missing compliances using the percentiles (*equipercentile equating assumption*)

2. Modelling strategy

- **EF: EFRON & FELDMAN (1991):** analysis of a randomized trial with partial non-compliance [Compliance as an Explanatory Variable in Clinical Trials, *JASA* 86, 9-17.]
- **FRANGAKIS & RUBIN (2002):** principal stratification
  - general framework to deal with non-compliance
  - earliest applications for all-or-none compliance (→ discrete strata)
- **JR: JIN & RUBIN (2008):** new analysis of Efron & Feldman data using *continuous principal strata* [Principal Stratification for Causal Inference With Extended Partial Compliance, *JASA* 103, 101-111.]
- **BARTOLUCCI & GRILLI (2010):** new analysis of Efron & Feldman data following the approach of Jin & Rubin but *with a different modeling strategy*
- Treatment indicator Z_i ($1=$drug, $0=$placebo)
- Potential outcomes (under *Strong access monotonicity*)
  - Compliance: d_i placebo, D_i drug
  - Outcome: $Y_i^{(0)}$ placebo, $Y_i^{(1)}$ drug
- Principal stratification
  - (d_i, D_i) principal strata (***continuous***)
  - $E(Y_i^{(1)} - Y_i^{(0)} \mid d_i, D_i)$ Principal Causal Effect (PCE)
- Regression models for the outcomes
  - $Y_i^{(0)}$ on d_i and D_i $\quad$ $Y_i^{(1)}$ on d_i and D_i

3. Copula for the joint distribution of the compliances

- d_i and D_i are never jointly observed but some information on their correlation is induced by the equations for the outcomes
- JR used a parametric specification consistent with *Negative side-effect monotonicity*: $D_i \le d_i$
- Critical issues:
  - How much information is available on the correlation between the compliances? How sensitive is the inference on the causal effect to the model on the compliances?
- We use a copula for the distribution of (d_i, D_i) A copula is a flexible way to define a joint distribution from the marginals
- We use a ***Plackett copula*** which has a single association parameter ψ
  - $0 < \psi < 1$ → negative association
  - $\psi = 1$ → independence
  - $\psi > 1$ → positive association
- Advantages of using a copula instead of a parametric density:
  - no constraints on the marginal distributions
  - association captured by a single parameter (to be estimated or used in a sensitivity analysis)

4. Model fitting via EM

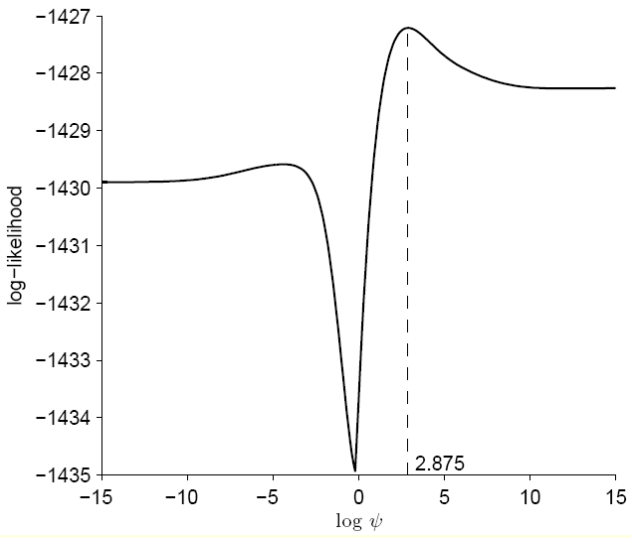
1. Compute the univariate empirical distribution functions of the two compliances d_i and D_i
2. For a set of values of the association parameter ψ
   i. Estimate the joint distribution function of (d_i, D_i) using the copula
   ii. Maximize the likelihood via EM
3. Plot the profile likelihood for ψ

This allows us to see how the different values of ψ are supported by the data and check for local maxima

- We begin with a general form with quadratic terms, interactions and heteroskedasticity and select via LR test
- FINAL MODEL FOR THE MEANS
- $E(Y_i^{(0)} | d_i, D_i) = -0.269 + 11.243 d_i$
- $E(Y_i^{(1)} | d_i, D_i) = -0.269 + 11.243 d_i - 21.878 D_i + 73.359 (d_i \times D_i)$
- **Principal Causal Effect:** $PCE(d_i, D_i) = (-21.878 + 73.359 d_i)\, D_i$
- The PCE depends on the dose of the drug D_i and the slope is
  - positive, except when $d_i < 0.298$ (but this is rare: 12.3% of the subjects in the placebo arm)
  - steeper at higher levels of the placebo compliance d_i

5. Maximum likelihood results

Profile log-likelihood for the Plackett association parameter

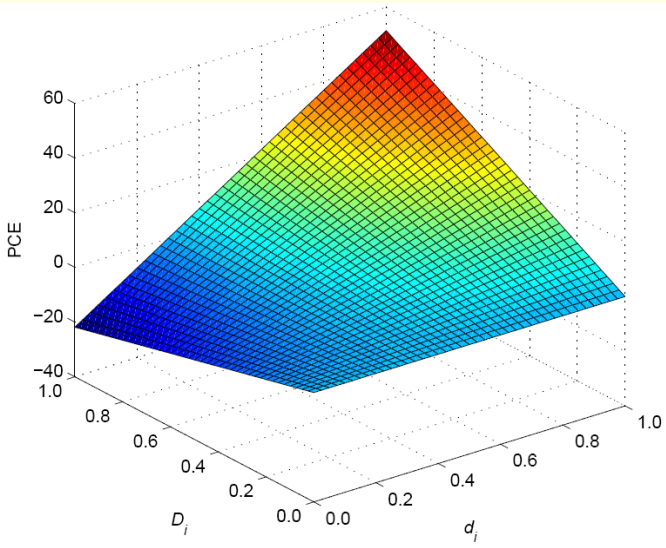


Point estimate of ψ is 17.727

Independence between d_i and D_i (i.e. ψ=1) is rejected (p-value<0.001)

Pearson correlation between d_i and D_i is 0.689

Principal Causal Effects (PCE) surface (ψ = ML estimate = 17.727)



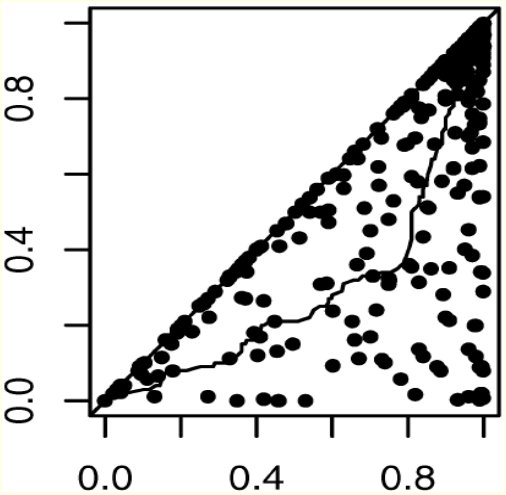
Good agreement with Jin and Rubin

At the median point (d_i=0.89, D_i=0.70) our ML estimate of PCE is 30.4 with bootstrap CI (22.5, 39.2)

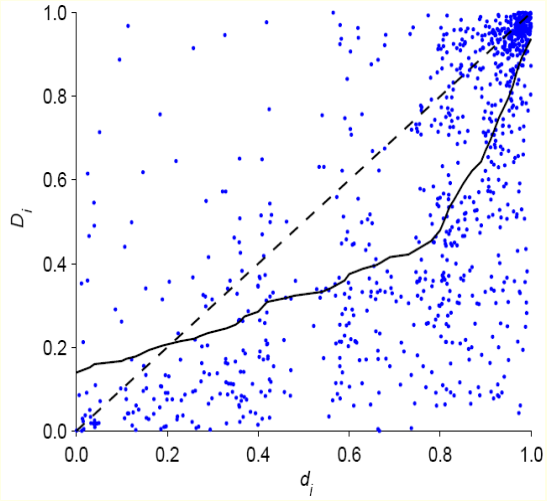
6. Association between the compliances

Random draws from the bivariate distribution of the compliances

Jin and Rubin (2008)



Bartolucci and Grilli (2010)



We relax the *negative side-effect monotonicity* (i.e. $D_i \leq d_i$)

→ 21.6% of the points go beyond the diagonal, corresponding to individuals with $D_i > d_i$

7. Sensitivity analysis

- The Plackett parameter ψ determining the association between drug and placebo compliances has *scarce empirical support* and it is identified thanks to the regression equations (which cannot be tested separately)
- The estimate of the PCE depends on the estimate of ψ : it is not advisable to base the inference exclusively on the point estimate of ψ
- We perform a **sensitivity analysis** to assess how the PCE depends on ψ (we let it vary in its profile likelihood interval):
  - at the median point (d_i=0.89, D_i=0.70): PCE $\in$(27.4, 34.8)
  - at the Q1-Q3 point (d_i=0.59, D_i=0.95): PCE $\in$(14.0, 29.5)
    - this couple of compliances is unlikely (far from the bulk of the data) → greater sensitivity, in addition to greater sampling variance: the bootstrap CI is very large (–10.8, 34.7)
- Principal Causal Effects are reliably estimated at drug and placebo compliance levels near the sample medians, while inference at unlikely compliance levels appears to be unduly affected by model assumptions