Using Regression Models to Analyze Randomized Trials: Asymptotically Valid Tests Despite Incorrect Regression Models

Michael Rosenblum (Berkeley/UCSF)

Mark J. van der Laan (Berkeley)

Motivation

- Regression models often used to analyze randomized trials.
- Models that adjust for baseline variables can add power to hypothesis tests.
- Danger: If model misspecified, may have large Type I and Type II error, even for large sample sizes.
- Research Question: For which models will hypothesis tests based on these models have asymptotically correct Type I error, even when the models are misspecified?

Outline

- Regression Models in Randomized Trials: Current Uses, Advantages, Limitations
- Example of a Hypothesis Test Based on Regression Model
- Related Work (Robins, Freedman)
- Our Results: For many regression models, certain simple hypothesis tests based on them have asymptotically correct Type I error, even when model misspecified. (Need to assume data I.I.D.)
- Open Problems

Models Often Used to Analyze Randomized Trials

Pocock et al. (2002) surveyed 50 clinical trial reports.

Findings: 36 used covariate adjustment

12 reports emphasized adjusted over unadjusted analysis.

"Nevertheless, the statistical emphasis on covariate adjustment is quite complex and often poorly understood, and there remains confusion as to what is an appropriate statistical strategy."

Advantages of Model-Based Tests

- Can have more power than Intention-to-Treat (ITT) based tests (e.g. if adjust for baseline variable(s) predictive of outcome).

  [Robinson and Jewell, 1991; Hernandez et al., 2004; Moore and van der Laan 2007; Freedman 2007]
- Can test for effect modification by baseline variables.

Misspecified Models Can Lead to Large Type I Error

- Robins (2004): for some classes of models, when the regression model is incorrectly specified, Type I error may be quite large even for large sample sizes.
- Potential for standard regression-based estimators to be asymptotically biased under the null hypothesis.
- Would lead to falsely rejecting null with probability tending to 1 as sample size tends to infinity (even with robust SE's).

Example of Model-Based Hypothesis Test in Rand. Trial

- Randomized trial of inhaled cyclosporine to prevent rejection after lung-transplantation. (Iacono et al. 2006)

  Outcome: number of severe rejection events per year of follow-up time.
- Some baseline variables known to be predictive of outcome: serologic mismatch, prior rejection event.
- Poisson Regression Used to Adjust for these.

Example (continued)

- Poisson model for conditional mean number of Rejection Events given Treatment (T), Serologic Mismatch (M) and Prior Rejection (P):

Log E(Rejections | T, M, P) =

$$\beta_0 + \beta_1 T + \beta_2 M + \beta_3 P$$

This Poisson model used to do hypothesis test:

If estimate of β_1 more than 1.96 SE's from 0, reject null hypothesis of no mean treatment effect within strata of M and P.

Example (continued)

Standard arguments to justify use of this Poisson model rely on assumption that it is correctly specified.

But what if this assumption is false?

Our main result implies that the above hypothesis test will have asymptotically correct Type I error, if the confidence interval is instead computed using a robust variance estimator (e.g. sandwich estimator), even when the model is misspecified.

Limitation of our results: we assume data I.I.D.

Null Hypothesis Being Tested

We test the null hypothesis of no mean treatment effect within strata of a set of baseline variables B.

That is, for T = treatment indicator,

E(Outcome | T =0, B) = E(Outcome | T=1, B).

This is a stronger (more restrictive) null hypothesis than no mean overall treatment effect:

E(Outcome| T=0) = E(Outcome | T=1).

It is a weaker (less restrictive) null hypothesis than no effect at all of treatment.

Related Work

D. Freedman (2007) shows that hypothesis tests based on ANCOVA model, that is, modeling E(Outcome| Treatment T, Baseline Variables B) by $\beta_0 + \beta_1 T + \beta_2 B,$ have asymptotically correct Type I error regardless of the data generating distribution.

J. Robins (2004) shows same for linear models with interaction terms. For example:

$$\beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \bullet B,$$

$$\beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \bullet B + \beta_4 B^2 + \beta_5 T \bullet B^2.$$

Scope of Our Results

- Our Results:

  -Apply to larger class of linear models than previously known.

  -Apply to large class of generalized linear models (including logistic regression, probit regression, Poisson regression).

  For example, the models

$$\text{logit}^{-1}(\beta_0 + \beta_1 T + \beta_2 B),$$

$$\Phi(\beta_0 + \beta_1 T + \beta_2 B_1 + \beta_3 B_2 + \beta_4 T \bullet B_1),$$

$$\exp(\beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \bullet B).$$

Summary of Our Main Result

- Null hypothesis we consider: No mean treatment effect within strata of baseline variables.
- Our Main Result:

  For a surprisingly large class of commonly used regression models, standard regression-based hypothesis tests (but using robust variance estimators) are guaranteed to have asymptotically correct Type I error, even when the models are incorrectly specified.

Hypothesis Testing Procedure

- Before looking at data:
  - Choose regression model satisfying constraints given in our paper (e.g. $\text{logit}^{-1}(\beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \bullet B)$).
  - Choose a coefficient β_i corresponding to a treatment term in the model (either β_1 or β_3 in example).
- Estimate the parameters of model using maximum likelihood estimation.
- Compute robust variance estimates with Huber sandwich estimator.
- Reject the null hypothesis of no mean treatment effect within strata of B if the estimate for β_i is more than 1.96 standard errors from 0.

Caveats of Hypothesis Testing Procedure

What if design matrix is not full rank?

What if maximum likelihood algorithm doesn't converge?

We always fail to reject the null hypothesis in these cases.

Since standard statistical software (e.g. R) will return a warning message when the design matrix is not full rank or when the maximum likelihood algorithm fails to converge, this condition is easy to check.

Limitations

- Assumption that data I.I.D.:

Not generally the case in randomized trial.

- Our results are asymptotic; performance not guaranteed for finite sample size
- Our results apply to hypothesis tests, not to estimation. For example, if hypothesis test rejects null, one cannot use same methods to create (asymptotically) valid confidence interval under the alternative.

Intuition behind Main Result

- When model misspecified, maximum likelihood estimator converges to maximizer β^* of expected log-likelihood $E_P \log p(X;\beta)$; for P the unknown data generating dist'n.
- Distribution corresponding to β^* can be viewed as projection of true data generating distribution onto model, based on Kullback-Leibler divergence.
- We prove that for certain models, components of β^* corresponding to treatment variable (T) are zero.

Models That Are Not Robust to Misspecification

- Models lacking "main terms," e.g.

$$\beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \bullet B + \beta_4 T \bullet B^2.$$

- Median Regression Models:

  Y = m(X, β) + ε

  for ε having Laplace distribution.

- More generally, models of the form:

  Y = m(X, β) + ε

  for ε mean 0, not normally distributed.

Effect Modification in Linear Models

- Our Results imply regression-based tests of effect modification are robust to model misspecification in certain settings:
  - Treatment T dichotomous,
  - Outcome Y is continuous,
  - Linear Model such as $\beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \bullet B,$
- Test whether baseline variable(s) B is effect modifier on additive scale: null hypothesis:

  E(Y|T=1,B) – E(Y|T=0,B) is constant.
- Reject null if estimate of β_3 more than 1.96 robust SE's from 0.

Regression Model vs. Semiparametric Model Based Tests

- Important work has been done using semiparametric methods to construct estimators and hypothesis tests that are robust to incorrectly specified models in randomized trials. [e.g. Robins, 1986; van der Laan and Robins, 2003; Tsiatis, 2006; Tsiatis et al., 2007; Zhang et al., 2007; Moore and van der Laan, 2007; Rubin and van der Laan, 2007].
- Our results use Regression methods:
  - Simpler to implement.
  - Can have more power if model approximately correctly specified.

Overall Recommendations

- Freedman (2008):

  First analyze experimental data following the ITT principle: compare rates or averages for subjects assigned to each treatment group.

  “This is simple, transparent, and fairly robust. Modeling should be secondary.”
- In model based tests, choose robust models and use robust variance estimators.

Open Problems

- Comparing Finite Sample Performance of Model Based Tests vs. Intention-to-Treat Based Tests. (This is what really matters in practice.)
- Proving results under framework that doesn't assume I.I.D. data, such as Neyman model used by Freedman (2007).

Thank you:

- David Freedman
- Estie Hudes (UCSF)
- Tor Neilands (UCSF)

Robust Variance Estimator

Huber's Sandwich estimator:

Let $l(\beta; Y, T, B) = \log p(Y \mid T, B, \beta)$ denote the log likelihood implied by regression model.

Let β^* denote the maximizer of $El(\beta; Y, T, B)$.

Let $\hat{\beta}$ denote the maximum likelihood estimator.

The asymptotic covariance of $\sqrt{n}(\hat{\beta} - \beta^*)$ is $\Sigma = B^{-1} W (B^{-1})^T$,

for $$B_{ij} = E \frac{\partial^2 l}{\partial \beta_i \partial \beta_j}, \qquad W_{ij} = E \frac{\partial l}{\partial \beta_i} \frac{\partial l}{\partial \beta_j}.$$

Models Having Robustness Property

I. Linear models for E(Outcome | T, B) of the form: $\sum \beta_j^{(0)} f_j(T,B) + \sum \beta_k^{(1)} g_k(B)$

where for every j, there is a k such that

$$E(f_j(T,B) \mid B) = g_k(B)$$

II. Generalized Linear Models with canonical links with linear parts of the form:

$$\sum \beta_j^{(0)} f_j(T) g_j(B) + \sum \beta_k^{(1)} g_k(B).$$

Model as Working Model

- Our approach is to never assume model is correct—we treat it as a "working model".
- Our goal is find simple tests based on regression models, that is, models of

E(Outcome | Treatment, Baseline Variables),

that have asymptotically correct Type I error regardless of the data generating distribution.

Advantage of such models over ITT is potentially more power.

Example of Linear Model with Robustness Property

For dichotomous treatment T
(taking values -1,1) and baseline variable B:

$$\beta_0 + \beta_1 \exp(B) + \beta_2 \exp(-B) + \beta_3 \exp(T \bullet B).$$