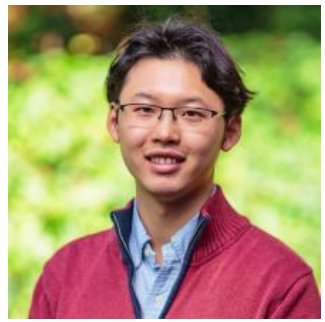
Non-Compliance & Instrumental Variables

Thomas Richardson

TUM Short Course Lecture II

Collaborators

Robin Evans
Oxford



F. Richard Guo
Berkeley
Cambridge

James M. Robins
Harvard

Outline

- Review:
  - Randomized experiments
  - Observational studies
- Instrumental variable model
  - Non-compliance
  - Combining observational studies
  - Bounds on the ACE; Testing the IV model
- Approaches to Statistical Inference
  - Bayesian: Naive vs. 'Transparent'
  - Frequentist approaches

Features and Caveats

Strengths:

- The approach will make minimal assumptions;
- Will be valid even if the instrument is ‘weak’.

Weaknesses:

- Only consider binary treatment and outcome;
- Will obtain bounds, not point identification.

Disclaimer: There is a huge literature on Instrumental Variables, here we focus on categorical treatment and outcome

Summary so far . . .

Randomized study

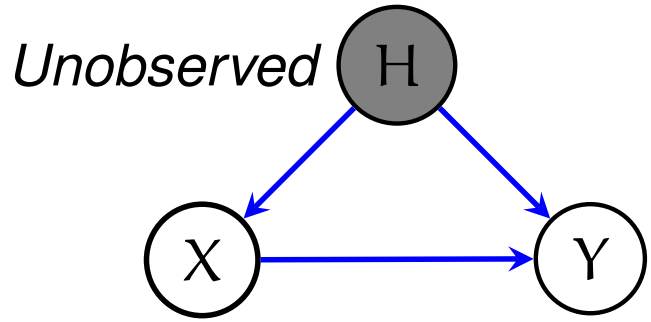
$X \perp\!\!\!\perp Y(x_0)$ & $X \perp\!\!\!\perp Y(x_1)$

$P(Y(x_0))$ and $P(Y(x_1))$ identified

$P(Y(x_0), Y(x_1))$ is a 1-d set

ACE identified

Observational study with confounding



$X \not\perp\!\!\!\perp Y(x_0)$ *or* $X \not\perp\!\!\!\perp Y(x_1)$

$P(Y(x_0))$, $P(Y(x_1))$ not identified

$P(Y(x_0), Y(x_1))$ is a 3-d set

ACE bounds of length 1

Instrumental Variable Models

Motivation: Non-Compliance

Oral ganciclovir fails to prevent CMV in HIV trial, *The Lancet*, (News Section), September 30, 1995

Oral ganciclovir has failed to prevent symptomatic cytomegalovirus (CMV) infections in a double-blind, placebo-controlled trial involving 994 patients with advanced HIV infection. The results of this US National Institute of Allergy and Infectious Diseases' Community Programs for Clinical Research on AIDS (CPCRA) study seems to contradict those of a similar trial indicating that oral ganciclovir significantly reduces CMV disease in HIV-infected patients. That study, conducted by the drug's manufacturer Syntex Research (now Roche Bioscience), was halted in 1994, after an interim analysis found that only 16 percent had developed CMV disease, compared with 30 percent in the placebo group.

In the CPCRA (protocol 023) study 662 volunteers were assigned to a treatment group and 332 to a placebo. Participants had to have CD4 cell counts of 100 or fewer per microliter of blood, evidence of CMV infection (by serology or culture) but no CMV disease. Patients in the treatment group took 3g ganciclovir daily. The primary endpoint of the trial was symptomatic CMV disease, defined as CMV retinal disease, diagnosed by ophthalmological examination, or CMV gastrointestinal mucosal disease, diagnosed on biopsy or on necropsy. Average follow-up was 15 months.

"Oral ganciclovir did not prevent symptomatic CMV disease to a clinically or statistically significant degree", the NIAID concluded in its Sept 18 announcement. Also, oral ganciclovir caused significantly more adverse effects than did placebo.

What the Lancet Didn't Say (I)

- The CPCRA data analysis used an "intention to treat" method
- Examines the differences between those asked to take Treatment vs. those asked to take Placebo;
  (Regardless of what they actually took.)

What the Lancet Didn't Say (II)

"After the CPCRA study began, the results of a different study involving 725 subjects became available which showed a 49% decrease in the number of clinical CMV infections in the group receiving prophylactic oral ganciclovir"

"*Consequently, for ethical reasons the CPCRA allowed the subjects in its placebo arm to take oral ganciclovir. The intention-to-treat analysis ignored this fact. The Lancet did not mention this problem*"

Hence the study compares the effect of asking people to take drug vs. placebo, but then giving everyone the drug anyway!

(In fact, the extent of bias is unclear since those in the control arm averaged 2.1 months of ganciclovir vs. 9.3 months in the treatment arm.)

Non-compliance: Cholestyramine trial

Z: assignment to treatment or control arm (randomized);
X: whether patient takes (more than certain amount of) drug;
Y: patient's health outcome.

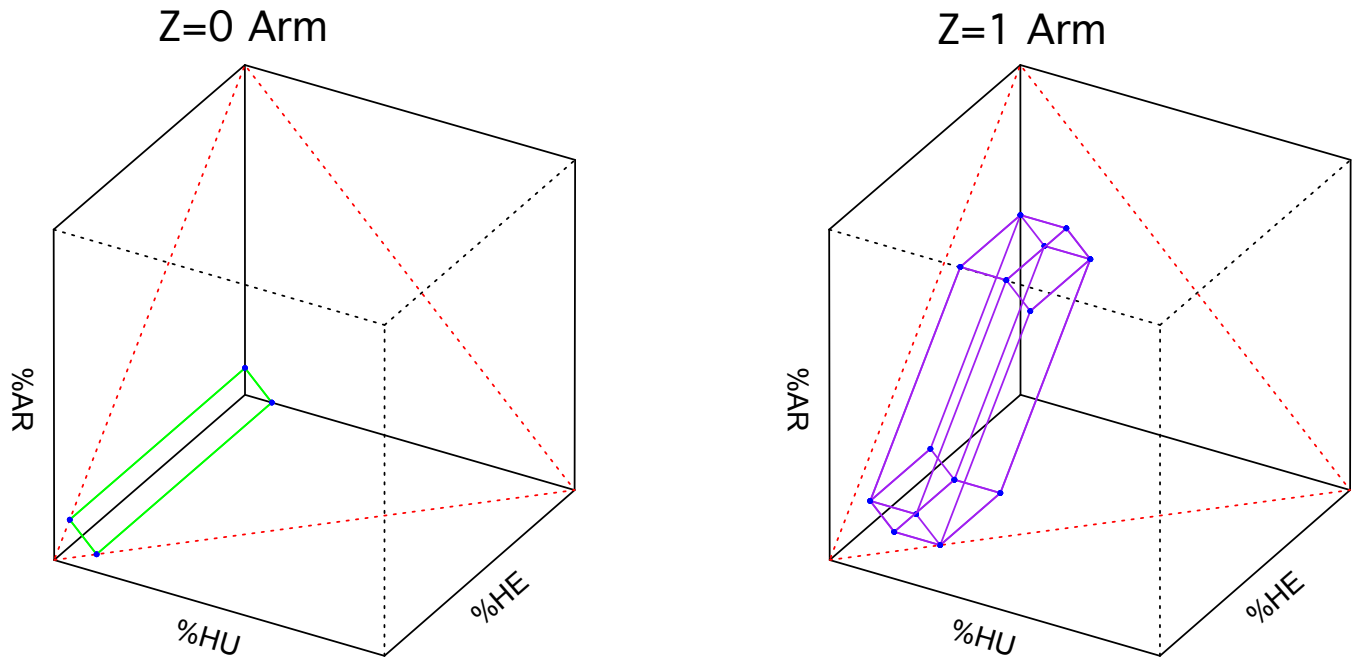
Z	X	Y	count
0	0	0	158
0	0	1	14
0	1	0	0
0	1	1	0
1	0	0	52
1	0	1	12
1	1	0	23
1	1	1	78

(Data originally considered by Efron and Feldman (1991); dichotomized by Pearl.)

We wish to find $\text{ACE}(X \to Y)$. Note $Z = 0 \;\Rightarrow\; X = 0$

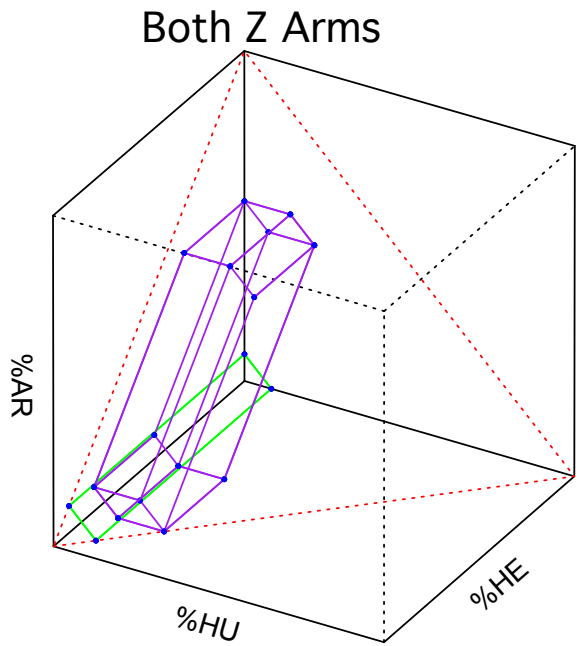
Idea: Analyze each Z arm as an observational study.

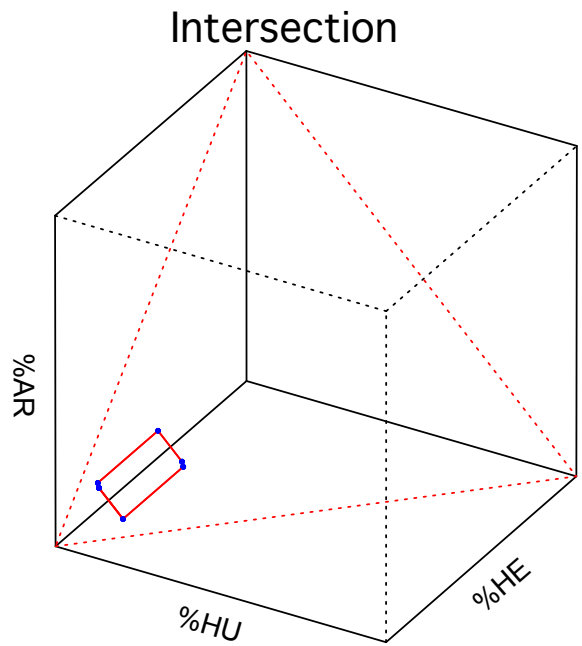
Each Z Arm



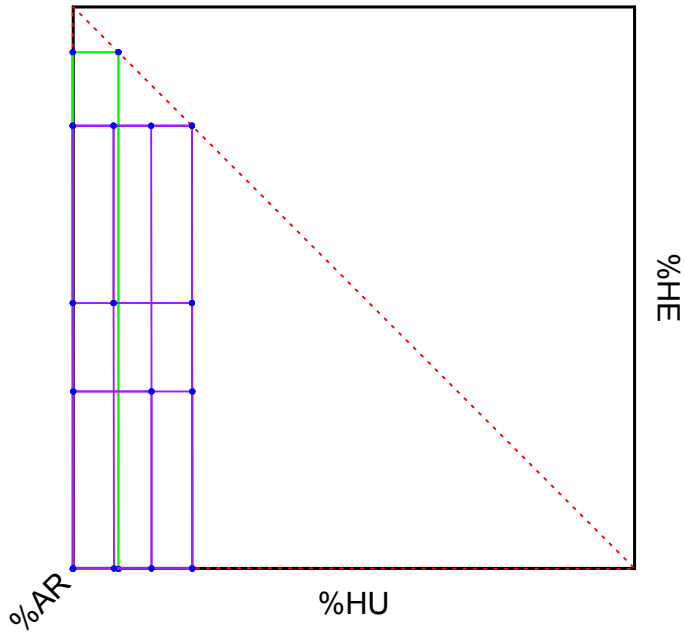
$Z = 0$ arm polytope is 2-d since $Z = 0 \Rightarrow X = 0$

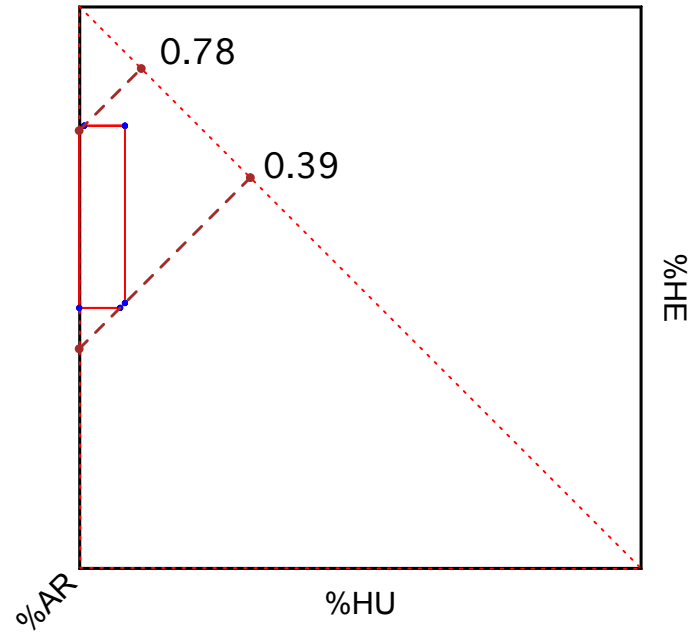
Combining the Arms





Obtaining ACE bounds



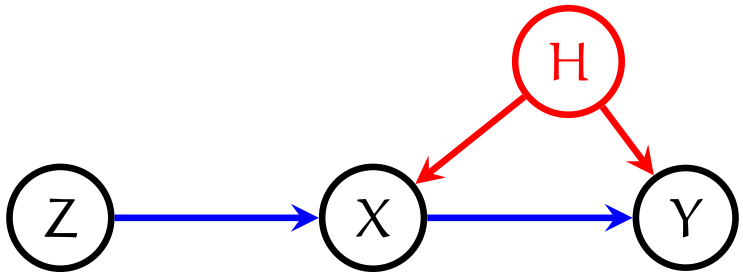


Upper bound is: 0.78; lower bound is 0.39

Note: ACE bounds for each arm contain 0, but not when combined. *Why?*

These are the Balke and Pearl (1993) bounds, obtained via prinicipal strata and computational algebra;

Binary Instrumental Variable Model



Z is said to be an 'instrument' for the effect of treatment X on outcome Y.

Potential outcomes: $X(z_0), X(z_1), Y(x_0), Y(x_1)$
here using x_i and z_i as shorthand for $x = i$ and $z = i$.

Randomization Assumption: for all x, z, $\quad Z \perp\!\!\!\perp \{X(z), Y(x)\}$;

Exclusion Assumption: No direct effect of Z on Y, so
$Y(x, z = 0) = Y(x, z = 1) \equiv Y(x)$;

Outcome Response Types

As before, in their response to the treatment, subjects belong to one of 4 response 'types':

$Y(x_0)$	$Y(x_1)$	Name
0	0	*Never Recover* (NR)
0	1	*Helped* (HE)
1	0	*Hurt* (HU)
1	1	*Always Recover* (AR)

Compliance Types

In their response to the instrument ("assigned treatment") patients are also of 4 types:

$X(z_0)$	$X(z_1)$	Name
1	1	*Always Taker (AT)*
0	0	*Never Taker (NT)*
0	1	*Complier (CO)*
1	0	*Defier (DE)*

Two-way table for binary IV model (Z binary)

	$p(x_0, y_0 \mid z_0)$	$p(x_0, y_1 \mid z_0)$	$p(x_1, y_0 \mid z_0)$	$p(x_1, y_1 \mid z_0)$
$p(x_0, y_0 \mid z_1)$	$p(\text{NT},\textit{NR}) + p(\text{NT},\textit{HE})$		$p(\text{DE},\textit{NR})$	$p(\text{DE},\textit{HE})$
$p(x_0, y_1 \mid z_1)$		$p(\text{NT},\textit{AR}) + p(\text{NT},\textit{HU})$	$p(\text{DE},\textit{HU})$	$p(\text{DE},\textit{AR})$
$p(x_1, y_0 \mid z_1)$	$p(\text{CO},\textit{NR})$	$p(\text{CO},\textit{HU})$	$p(\text{AT},\textit{NR}) + p(\text{AT},\textit{HU})$	
$p(x_1, y_1 \mid z_1)$	$p(\text{CO},\textit{HE})$	$p(\text{CO},\textit{AR})$		$p(\text{AT},\textit{AR}) + p(\text{AT},\textit{HE})$

Bounds on ACE: Previous results

- Robins (1989) and Manski (1990) derived the 'natural' bounds:

$$p(y_1 \mid z_1) - p(y_1 \mid z_0) - p(y_1, x_0 \mid z_1) - p(y_0, x_1 \mid z_0)$$

$$\leqslant \quad ACE(X \to Y) \quad \leqslant$$

$$p(y_1 \mid z_1) - p(y_1 \mid z_0) + p(y_0, x_0 \mid z_1) + p(y_1, x_1 \mid z_0)$$

  These are not sharp if there are Defiers;

  they follow from (and are sharp) under $Z \perp\!\!\!\perp Y(x_0)$, $Z \perp\!\!\!\perp Y(x_1)$.

- Balke and Pearl (1993) derived closed form expression for ACE bounds via computational algebra;

  Resulting expressions are maxima and minima over 8 different expressions.

- Dawid (2002) re-derived these bounds without (explicitly!) using potential outcomes, again using computational algebra.

An advantage of the approach taken here is that it extends to Z with finitely many levels.

Extending to p levels of Z: Identification

Solving the identification problem requires no further work:
Intersect p polytopes.

The set of distributions $P(Y(x_0), Y(x_1))$ that are compatible with $P(X, Y|Z)$ is still given by the following six pairs of parallel planes:

$$
\begin{aligned}
0 \leqslant \quad & \%NR \quad && \leqslant \min_i p(y_0 \mid z_i) \\
0 \leqslant \quad & \%AR \quad && \leqslant \min_i p(y_1 \mid z_i) \\
0 \leqslant \quad & \%HE \quad && \leqslant \min_i \{p(x_1, y_1 \mid z_i) + p(x_0, y_0 \mid z_i)\} \\
0 \leqslant \quad & \%HU \quad && \leqslant \min_i \{p(x_0, y_1 \mid z_i) + p(x_1, y_0 \mid z_i)\} \\
\max_i p(x_0, y_1 \mid z_i) \leqslant \quad & P(Y(x_0)=1) \quad && \leqslant \min_i \{1 - p(x_0, y_0 \mid z_i)\} \\
\max_i p(x_1, y_1 \mid z_i) \leqslant \quad & P(Y(x_1)=1) \quad && \leqslant \min_i \{1 - p(x_1, y_0 \mid z_i)\}
\end{aligned}
$$

Generalizing to more levels of Z

Theorem (R+Robins 2014)

If X, Y binary and Z with p levels then under the IV model:

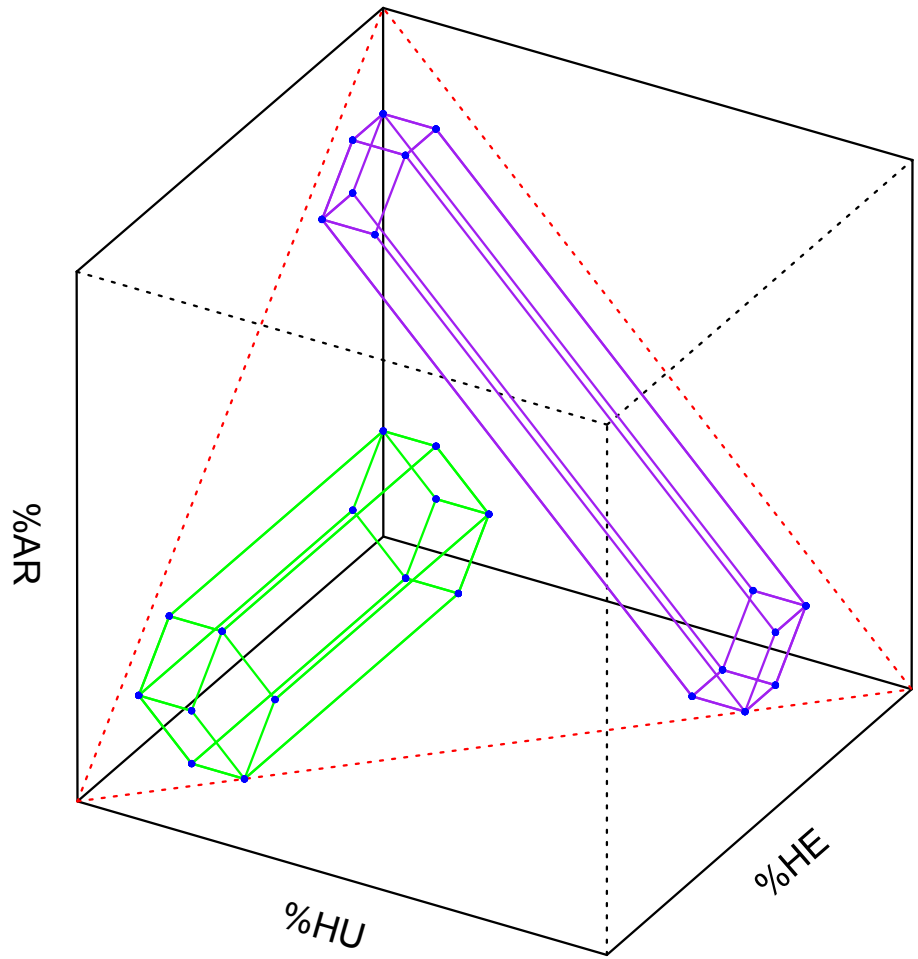
$$1 - g(1,0) - g(0,1) \leqslant ACE(X \to Y) \leqslant g(0,0) + g(1,1) - 1,$$

where

$$g(i,j) \equiv \min \left\{ \min_z \left[P(X{=}i, Y{=}j \mid Z{=}z) + P(X{=}1-i \mid Z{=}z) \right], \right.$$
$$\left. \min_{z,\tilde{z}:\, z \neq \tilde{z}} \left[P(X{=}i, Y{=}j \mid Z{=}z) + P(X{=}1-i, Y{=}0 \mid Z{=}z) \right. \right.$$
$$\left. \left. + \quad P(X{=}i, Y{=}j \mid Z{=}\tilde{z}) + P(X{=}1-i, Y{=}1 \mid Z{=}\tilde{z}) \right] \right\}.$$

This exploits the fact that $P(Y(x_0) = 1)$ and $P(Y(x_1) = 1)$ are variation independent.

Polytopes may not intersect



$\Rightarrow$ Model places testable constraints on $P(X, Y \mid Z)$.

Model for observables

For Z binary requiring that the polytopes intersect leads to the following:
If $p(X, Y \mid Z)$ is compatible with the binary IV model iff

$$\begin{aligned}
p(Y{=}0, X{=}0 \mid Z{=}0) + p(Y{=}1, X{=}0 \mid Z{=}1) &\leqslant 1, \\
p(Y{=}0, X{=}0 \mid Z{=}1) + p(Y{=}1, X{=}0 \mid Z{=}0) &\leqslant 1, \\
p(Y{=}0, X{=}1 \mid Z{=}0) + p(Y{=}1, X{=}1 \mid Z{=}1) &\leqslant 1, \\
p(Y{=}0, X{=}1 \mid Z{=}1) + p(Y{=}1, X{=}1 \mid Z{=}0) &\leqslant 1,
\end{aligned}$$

This describes a subset of $\Delta^3 \times \Delta^3$.

These are the IV inequalities of Pearl (1995) and Bonet (2001);
they provide a *falsification* test of the binary IV model.

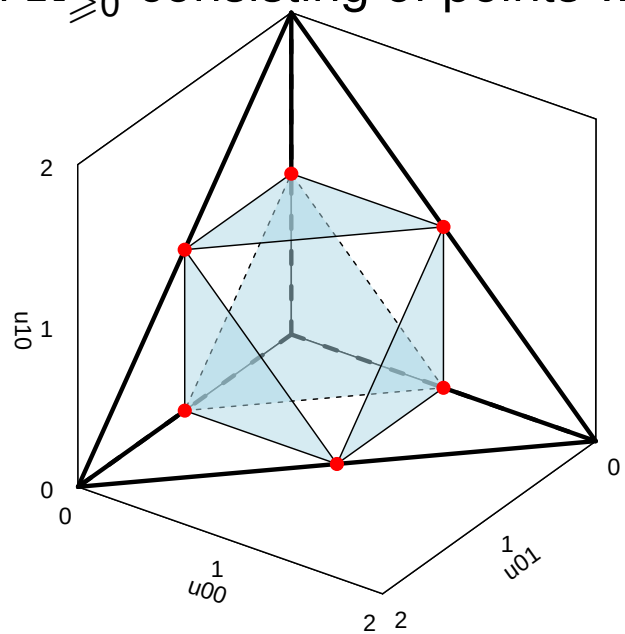
Can be interpreted as bounding away from zero $E[Y(z_1, x) - Y(z_0, x)]$,
the average direct effect of Z on Y, holding X fixed at x.

Visualizing the restrictions

Define the following variables:

$$
\begin{aligned}
u_{00} &\equiv p(Y=0, X=0 \mid Z=0) + p(Y=1, X=0 \mid Z=1) &\leqslant 1,\\
u_{01} &\equiv p(Y=0, X=0 \mid Z=1) + p(Y=1, X=0 \mid Z=0) &\leqslant 1,\\
u_{10} &\equiv p(Y=0, X=1 \mid Z=0) + p(Y=1, X=1 \mid Z=1) &\leqslant 1,\\
u_{11} &\equiv p(Y=0, X=1 \mid Z=1) + p(Y=1, X=1 \mid Z=0) &\leqslant 1,
\end{aligned}
$$

Since $u_{00} + u_{01} + u_{10} + u_{11} = 2$ these variables
live in a 3-d simplex of $\mathbb{R}^4_{\geqslant 0}$ consisting of points with sum $= 2$.



It follows that at most one inequality can be violated (see also Cai, Kuroki, Pearl, Tian, 2008).

Extensions and related work

Lauritzen & Ramsahai (2011) provide an approach for testing these restrictions via a likelihood procedure combined with a bootstrap due to Andrews 2000.

Wang, Robins, R (2016) describe a procedure that allows to test the restriction of the binary model in the presence of baseline covariates based on the Gail-Simon test.

With more levels of Z there are additional constraints that involve probabilities from up to 4 arms; see Bonet (2001); Kédagni, Mourifié (2019).

Kédagni and Mourifié also provide a testing procedure allowing for continuous Z using the Sample Splitting Procedure of Chernozhukov et al. (2015).

Bayesian Inference

Bayesian Inference (naive approach)

What about sampling variability?

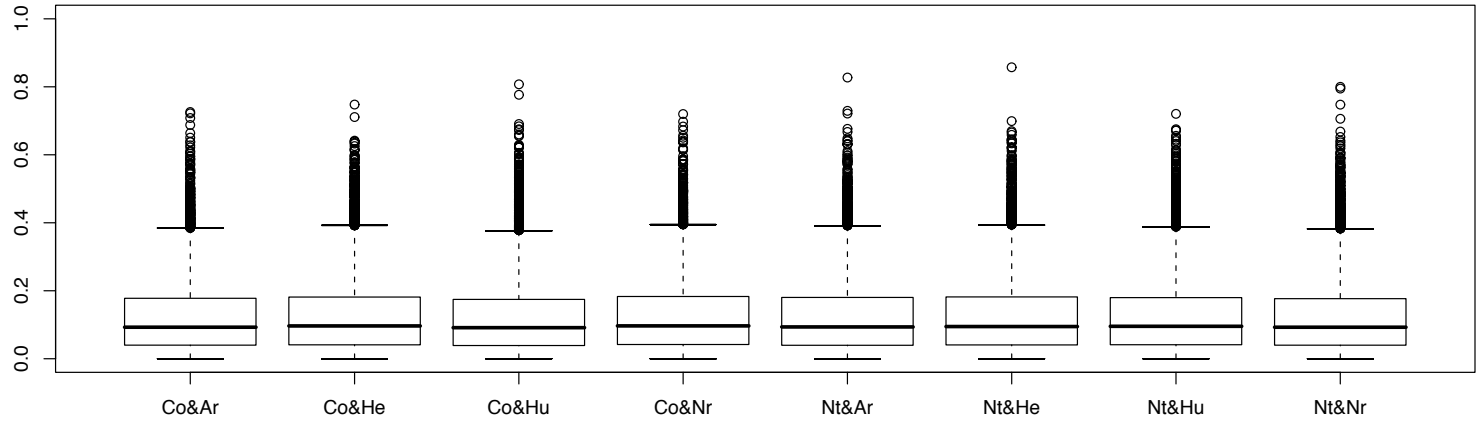
Bounds were obtained by just plugging in the empirical $\hat{p}(x, y \mid z)$ as if it were the truth.

$\Rightarrow$ Simple Bayesian approach (Pearl, 2000, Ch.8): put a (Dirichlet) prior on the set of distributions:

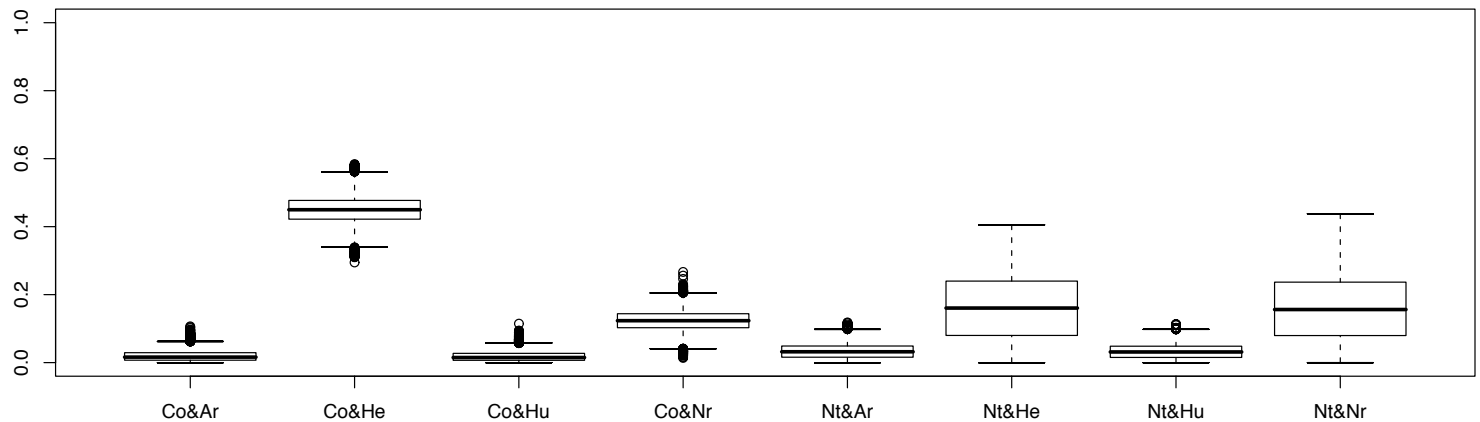
$$\{P(X(z_0), X(z_1), Y(x_0), Y(x_1))\}.$$

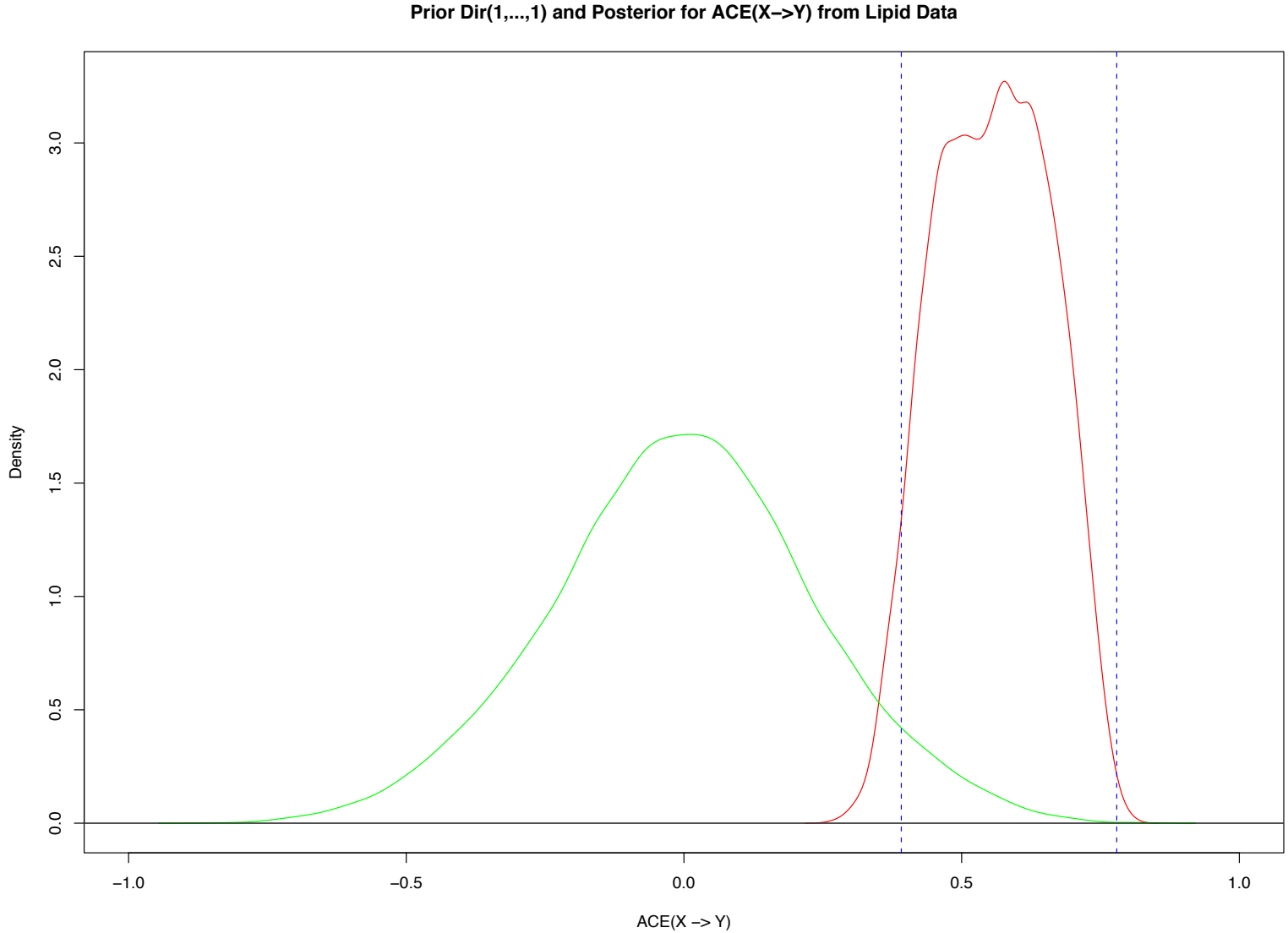
$\Rightarrow$ Use MCMC to sample from posterior distribution.

Marginal Prior distributions; Dir(1,...,1)

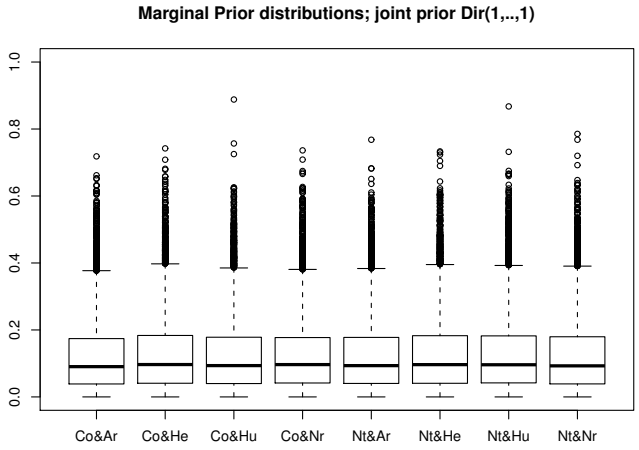


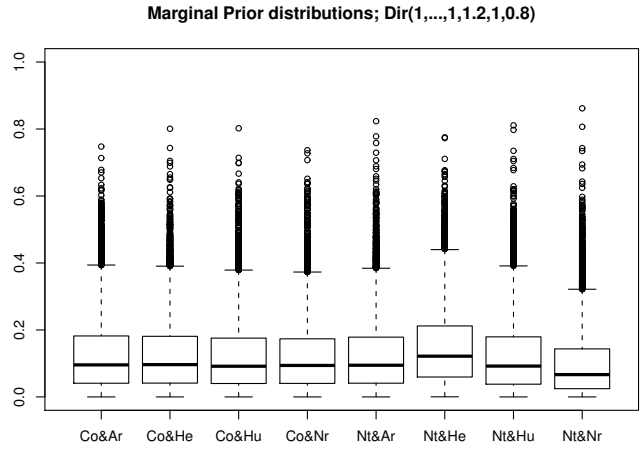
Marginal Posterior distributions from Lipid Data; Prior Dir(1,...,1)

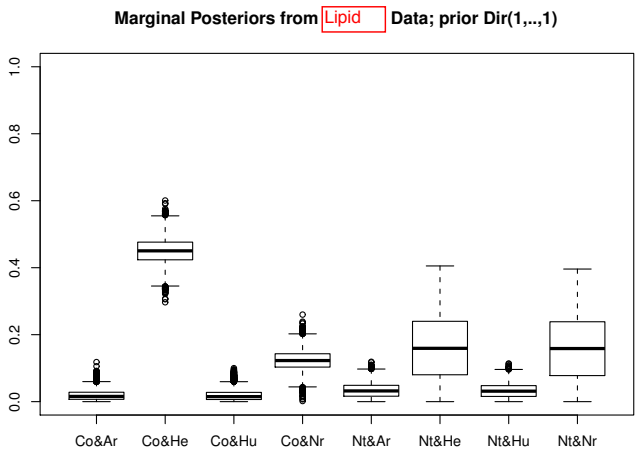


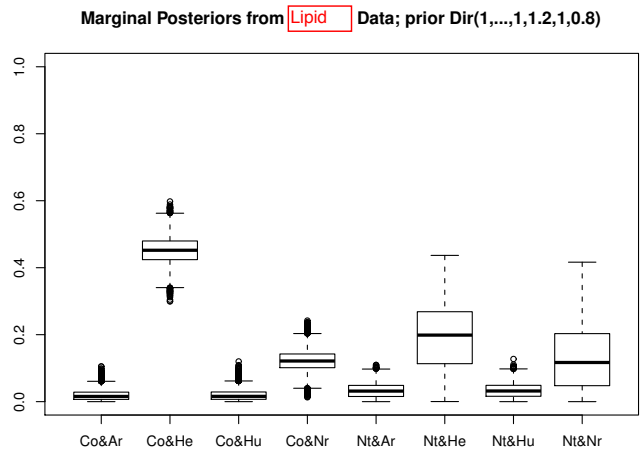


Prior (green); Posterior (red)


Marginal Prior distributions; joint prior Dir(1,..,1)
0.0
0.2
0.4
0.6
0.8
1.0
Co&Ar
Co&He
Co&Hu
Co&Nr
Nt&Ar
Nt&He
Nt&Hu
Nt&Nr

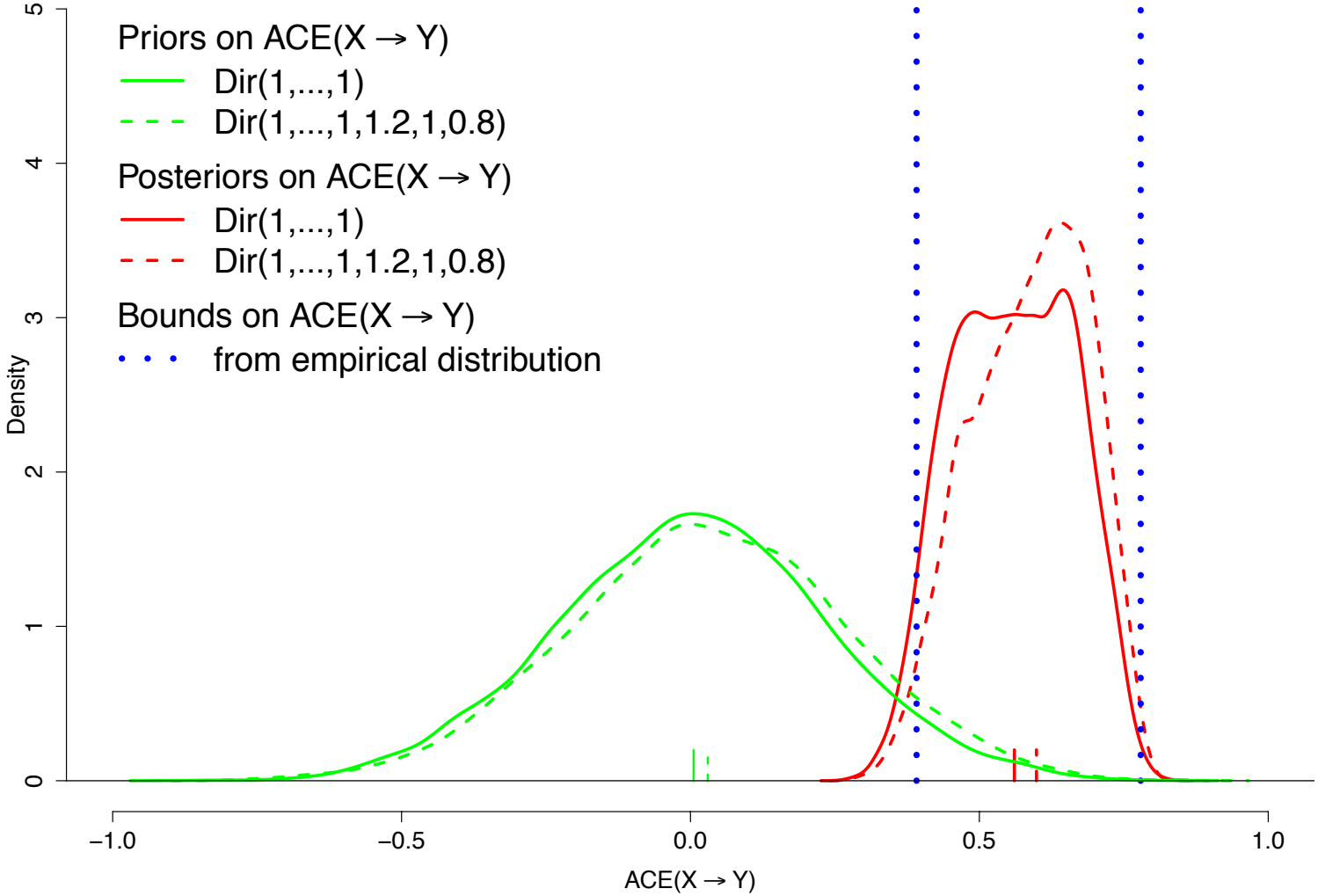

Marginal Prior distributions; Dir(1,...,1,1.2,1,0.8)
0.0
0.2
0.4
0.6
0.8
1.0
Co&Ar
Co&He
Co&Hu
Co&Nr
Nt&Ar
Nt&He
Nt&Hu
Nt&Nr


Marginal Posteriors from Lipid Data; prior Dir(1,..,1)
0.0
0.2
0.4
0.6
0.8
1.0
Co&Ar
Co&He
Co&Hu
Co&Nr
Nt&Ar
Nt&He
Nt&Hu
Nt&Nr


Marginal Posteriors from Lipid Data; prior Dir(1,...,1,1.2,1,0.8)
0.0
0.2
0.4
0.6
0.8
1.0
Co&Ar
Co&He
Co&Hu
Co&Nr
Nt&Ar
Nt&He
Nt&Hu
Nt&Nr

Prior and posterior on ACE(X → Y) for Lipid data

Uniform & perturbed uniform priors on potential outcomes



Is the problem caused by the priors not being diffuse enough?

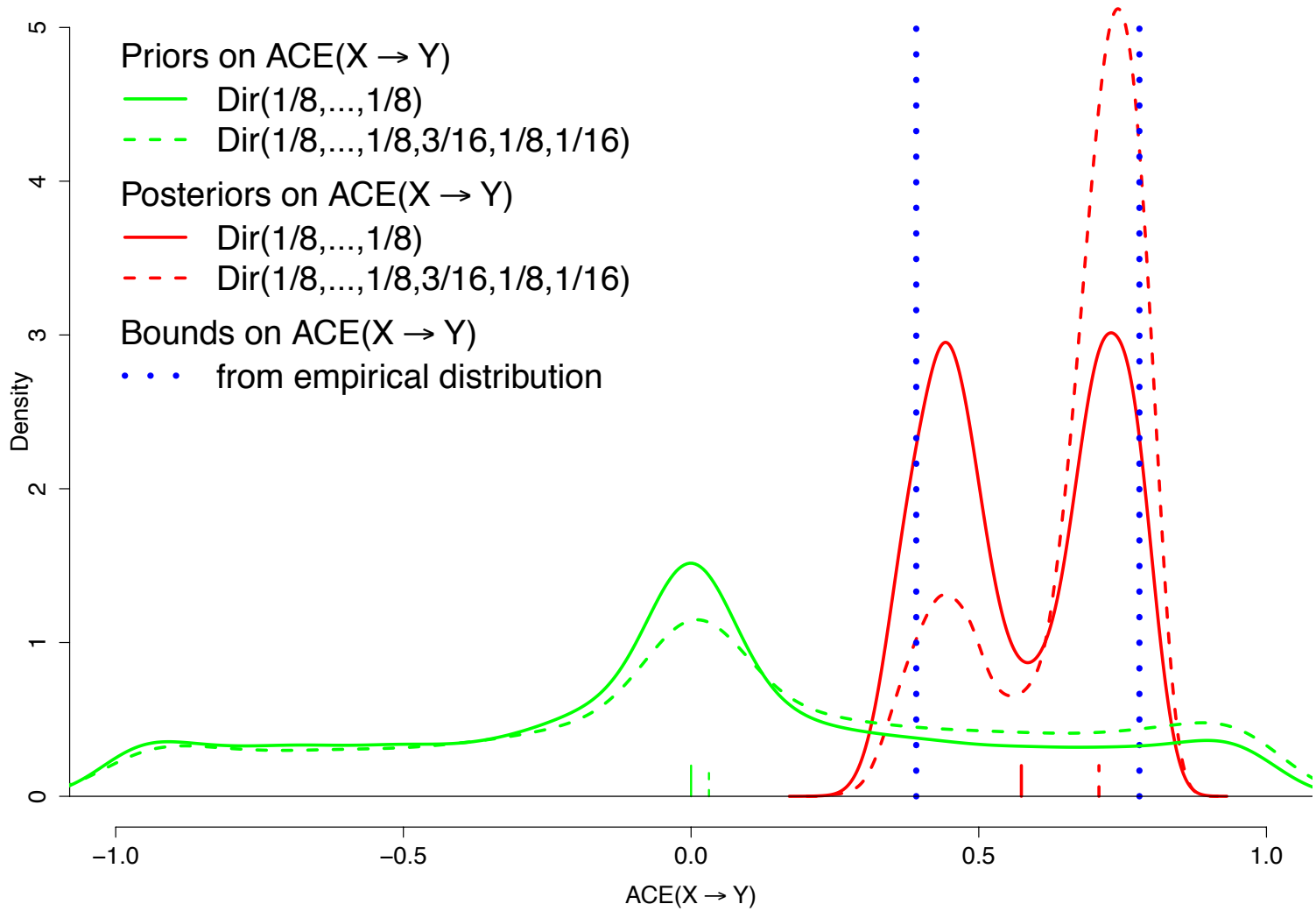
Try a 'unit' information prior:

$$p(\{X(z), Y(x)\}) \sim \text{Dir}(1/8, \ldots, 1/8)$$

vs.

$$p(\{X(z), Y(x)\}) \sim \text{Dir}(1/8, \ldots, 1/8, 3/16, 1/8, 1/16)$$

(?Though 'unit' of whose information?)

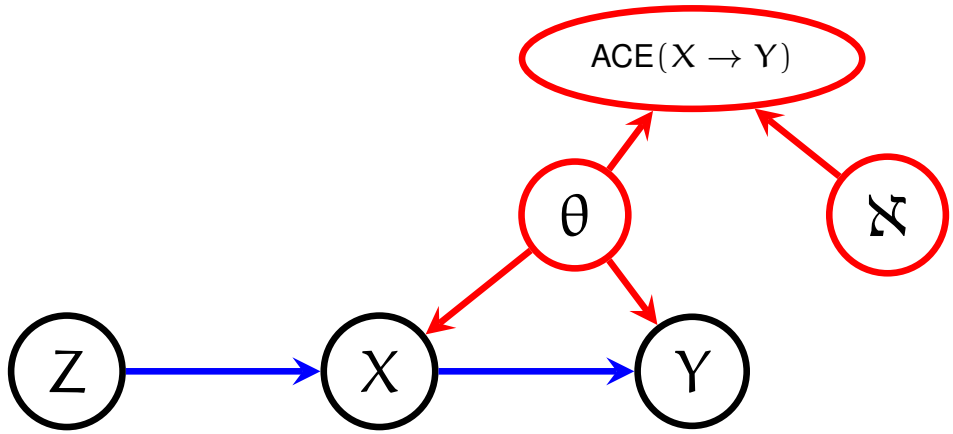


Basic problem: 16 types (= 15 params) but data is only 6 dimensional.

Two-way table for binary IV model (Z binary)

	$p(x_0, y_0 \mid z_0)$	$p(x_0, y_1 \mid z_0)$	$p(x_1, y_0 \mid z_0)$	$p(x_1, y_1 \mid z_0)$
$p(x_0, y_0 \mid z_1)$	$p(\text{NT},\textit{NR}) + p(\text{NT},\textit{HE})$		$p(\text{DE},\textit{NR})$	$p(\text{DE},\textit{HE})$
$p(x_0, y_1 \mid z_1)$		$p(\text{NT},\textit{AR}) + p(\text{NT},\textit{HU})$	$p(\text{DE},\textit{HU})$	$p(\text{DE},\textit{AR})$
$p(x_1, y_0 \mid z_1)$	$p(\text{CO},\textit{NR})$	$p(\text{CO},\textit{HU})$	$p(\text{AT},\textit{NR}) + p(\text{AT},\textit{HU})$	
$p(x_1, y_1 \mid z_1)$	$p(\text{CO},\textit{HE})$	$p(\text{CO},\textit{AR})$		$p(\text{AT},\textit{AR}) + p(\text{AT},\textit{HE})$

Better Bayesian approach: transparent re-parametrization



We will re-parameterize:

$$\{P(X(z_0), X(z_1), Y(x_0), Y(x_1))\} \quad \leftrightarrow \quad (\theta, \aleph)$$

θ is a 6 dim. parameter, (completely!) identifiable from $P(X, Y \mid Z)$.
$\aleph$ is a 9 dim. parameter, (completely!) non-identifiable.

$$P(\theta, \aleph) = P(\theta)P(\aleph)$$
$$P(\theta, \aleph \mid Z, X, Y) = P(\theta|Z, X, Y)p(\aleph)$$

Note that $\aleph \perp\!\!\!\perp Z, X, Y$

Simple implementation

Recall that the binary IV model is defined by the inequalities:

$$\begin{aligned}
P(Y=0, X=0 \mid Z=0) + P(Y=1, X=0 \mid Z=1) &\leqslant 1, \\
P(Y=0, X=0 \mid Z=1) + P(Y=1, X=0 \mid Z=0) &\leqslant 1, \\
P(Y=0, X=1 \mid Z=0) + P(Y=1, X=1 \mid Z=1) &\leqslant 1, \\
P(Y=0, X=1 \mid Z=1) + P(Y=1, X=1 \mid Z=0) &\leqslant 1.
\end{aligned} \tag{1}$$

Prior: Dirichlet on $P(X, Y \mid Z)$ restricted (and re-normalized) to those distributions obeying (1).

Posterior: The usual Dirichlet posterior, again restricted to those distributions obeying (1).

Inference may be performed by 'straight' Monte-Carlo.

Bayesian Monte Carlo Algorithm in Detail

To obtain the posterior distribution on the ACE bounds, perform the following steps:

1. Specify Dirichlet $(\alpha_{00z}, \alpha_{01z}, \alpha_{10z}, \alpha_{11z})$ priors on $p(x, y|z)$ for $z = 0, 1$.
2. Compute the posteriors in the usual way:
   Dirichlet $(\alpha_{00z} + n_{00z}, \alpha_{01z} + n_{01z}, \alpha_{10z} + n_{10z}, \alpha_{11z} + n_{11z})$
   where n_{ijz} is the number of observations with $X = i$, $Y = j$, $Z = z$.
3. Simulate $p^{(1)}(x, y|z), \ldots, p^{(N)}(x, y|z)$ from this posterior.
4. Throw out any $p^{(i)}(x, y|z)$ violating the inequalities (1).
5. Compute upper and lower bounds on the ACE from each distribution $p^{(i)}(x, y|z)$ remaining after step 4.

Back to Lipid Data...

Parametrize IV model as:

$$p(X, Y \mid Z) \text{ obeying IV inequalities}$$

Priors:

Dir$(1, 1)$ for $p(x = 0, y \mid z = 0)$, since $p(x=1, y \mid z=0) = 0$, (no AT or DE)

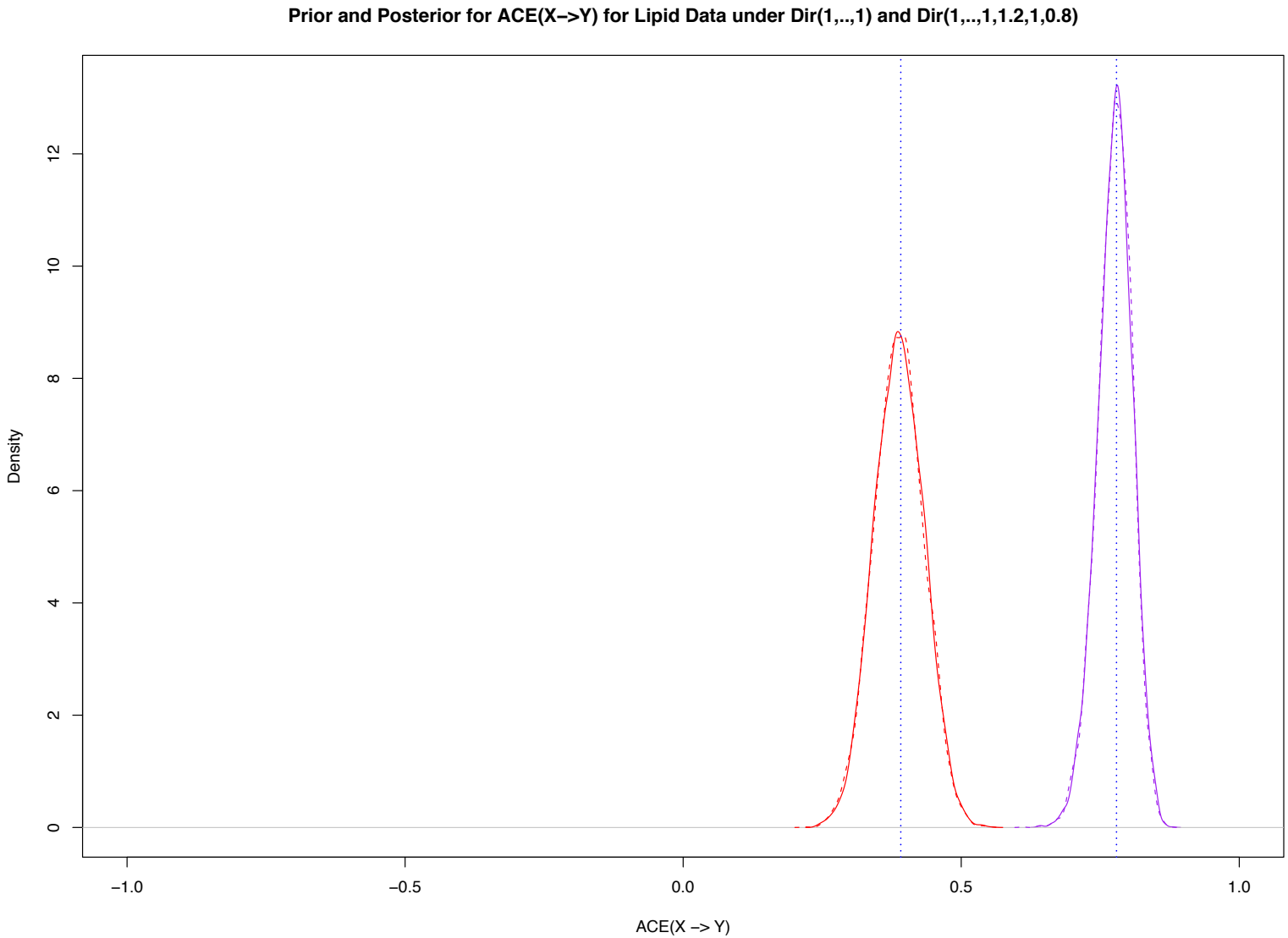
Dir$(1, 1, 1, 1)$ for $p(x, y \mid z = 1)$.

We then truncate and renormalize using the IV inequalities (1).

Model test:

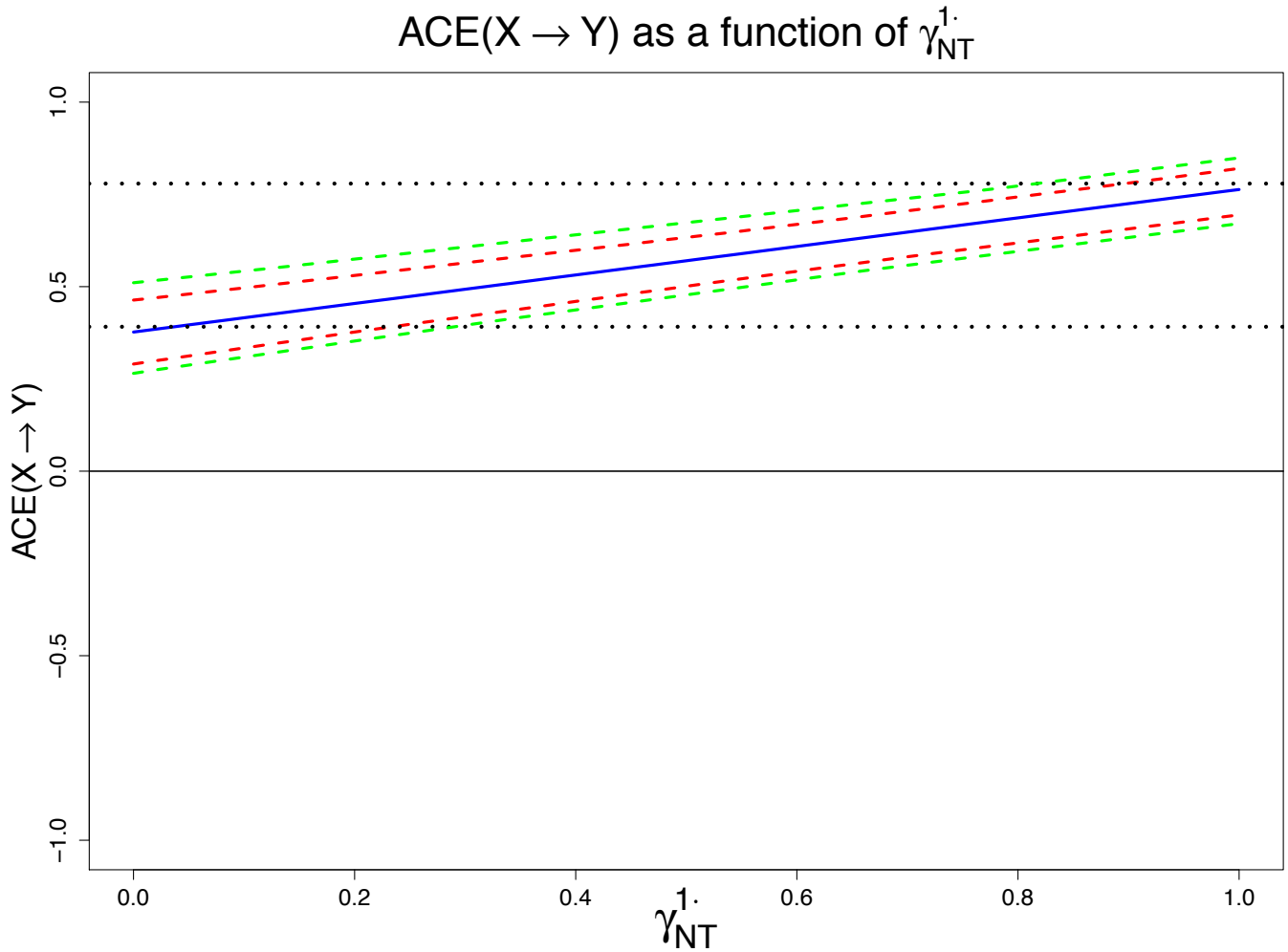
Given uniform prior on the whole simplex, prior probability was 0.5;
Posterior probability that IV model holds: 0.632.

Bayesian analysis on Bounds



Perturbing the prior on the observables has very little effect.

Sensitivity Analysis



Here $\gamma^{1\cdot}_{NT} \equiv P(Y(x_1) = 1 \mid NT)$, the probability of a good outcome for Never Takers, if they were to get the drug; this is completely unidentified.

Frequentist Inference for ACE after testing the IV model

$$
\begin{aligned}
P(Y=0, X=0 \mid Z=0) + P(Y=1, X=0 \mid Z=1) &\leqslant 1, \\
P(Y=0, X=0 \mid Z=1) + P(Y=1, X=0 \mid Z=0) &\leqslant 1, \\
P(Y=0, X=1 \mid Z=0) + P(Y=1, X=1 \mid Z=1) &\leqslant 1, \\
P(Y=0, X=1 \mid Z=1) + P(Y=1, X=1 \mid Z=0) &\leqslant 1.
\end{aligned}
\tag{1}
$$

- Inference via asymptotic distribution of likelihood ratio test? (Lauritzen & Ramsahai, 2011)
  - *But asymptotic distribution is not invariant due to boundaries in the model...*
  - "$np_i > 5$" ?
- Bootstrap?
  - *Also fails on the boundary* (Andrews 2000)
- Sample splitting?
  - *Separate subsample for each inequality tested.*

Alternative Simple Approach

1. Construct a joint confidence region for the distributions $\{p(X, Y \mid z)$ for each arm $z\}$;
2. Intersect this joint region with the model defining inequalities;
3. Compute bounds on the ACE over the region found in 2.

Subproblem:

Desired Properties for Confidence Region for Multinomial:

- Non-asymptotic;
- Convex;
- "Small" $\Rightarrow$ Approaching the asymptotic region in large samples;
- Easily Computable.

Confidence Region from Tail-Bound on KL-divergence

Data:

$$(X_1, \ldots, X_k) \sim \text{Mult}(n; (p_1, \ldots, p_k)), \tag{2}$$

Find a critical value t_α such that for *all* $p \in \Delta^{k-1}$:

$$\mathbb{P}\left(n\, \mathcal{D}(\widehat{p} \| p) > t_\alpha\right) \leqslant \alpha. \tag{3}$$

where $(\widehat{p}_1, \ldots, \widehat{p}_k) \equiv (X_1/n, \ldots, X_k/n)$.

$(1-\alpha)$% Confidence Region: $R \equiv \{p : n\, \mathcal{D}(\widehat{p} \| p) \leqslant t_\alpha\}$.
The resulting region is convex.

Chernoff's method

Bound the MGF of $n\,\mathcal{D}(\hat{p}\|p))$:

$$\varphi_{k,n}(\lambda, p) \equiv \mathbb{E}_p \exp\left(\lambda n\,\mathcal{D}(\widehat{p}\|p)\right), \tag{4}$$

Now:

$$\begin{aligned}\varphi_{k,n}(\lambda, p) &= \sum_{x_1,\ldots,x_k} \binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} p_i^{x_i} \left\{ \frac{\binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} \widehat{p}_i^{x_i}}{\binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} p_i^{x_i}} \right\}^{\lambda} \\ &= \sum_{x_1,\ldots,x_k} \binom{n}{x_1,\ldots,x_k} \left\{ \prod_{i=1}^{k} \widehat{p}_i^{x_i} \right\}^{\lambda} \left\{ \prod_{i=1}^{k} p_i^{x_i} \right\}^{1-\lambda},\end{aligned} \tag{5}$$

Chernoff's method

Bound the MGF of $n\,\mathcal{D}(\hat{p}\|p))$:

$$\varphi_{k,n}(\lambda, p) \equiv \mathbb{E}_p \exp\left(\lambda n\,\mathcal{D}(\widehat{p}\|p)\right), \tag{6}$$

Now:

$$\begin{aligned}
\varphi_{k,n}(\lambda, p) &= \sum_{x_1,\ldots,x_k} \binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} p_i^{x_i} \left\{ \frac{\binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} \widehat{p}_i^{x_i}}{\binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} p_i^{x_i}} \right\}^{\lambda} \\
&= \sum_{x_1,\ldots,x_k} \binom{n}{x_1,\ldots,x_k} \left\{ \prod_{i=1}^{k} \widehat{p}_i^{x_i} \right\}^{\lambda} \left\{ \prod_{i=1}^{k} p_i^{x_i} \right\}^{1-\lambda}, \\
&\leqslant \sum_{x_1,\ldots,x_k} \binom{n}{x_1,\ldots,x_k} \prod_{i=1}^{k} [\lambda \widehat{p}_i + (1-\lambda) p_i]^{x_i}
\end{aligned} \tag{7}$$

This is an upper bound, but it appears to depend on the (unknown) p.

Happy Fact!

The dependence is *illusory*:

$$G_{k,n}(\lambda, p) \equiv \sum_{x_1,\ldots,x_k} \binom{n}{x_1, \ldots, x_k} \prod_{i=1}^{k} [\lambda x_i/n + (1-\lambda)p_i]^{x_i}, \tag{8}$$

does *not* depend on p (!)

First noted in the case $k = 2$ by Rohit Agrawal (2019); General case: Guo & R (2020).

General Expression

Theorem (Guo & R 2020)

For $k \geqslant 2$, $n \geqslant 0$ and $0 \leqslant \lambda \leqslant 1$, it holds that

$$G_{k,n}(\lambda) = \sum_{m=0}^{n} \frac{n!}{n^m (n-m)!} \binom{m+k-2}{k-2} \lambda^m. \tag{9}$$

These are related to Charlier and Laguerre polynomials.

This bound on the MGF $\varphi_{k,n}(\lambda, p)$ is asymptotically tight in the large deviation sense.

Optimizing over λ

For any $\lambda \in [0, 1]$, we have

$$\mathbb{P}\left(n\,\mathcal{D}(\hat{p}_{k,n}\|p) > t\right) \leqslant \exp(-\lambda t) G_{k,n}(\lambda). \tag{10}$$

Can obtain the tightest bound by minimizing over $\lambda \in [0, 1]$.
$\Rightarrow$ In general a 1-dimensional optimization (albeit non-convex).

Other choices for λ

- (First idea) Use the value of λ that is optimal for $n \to \infty$:

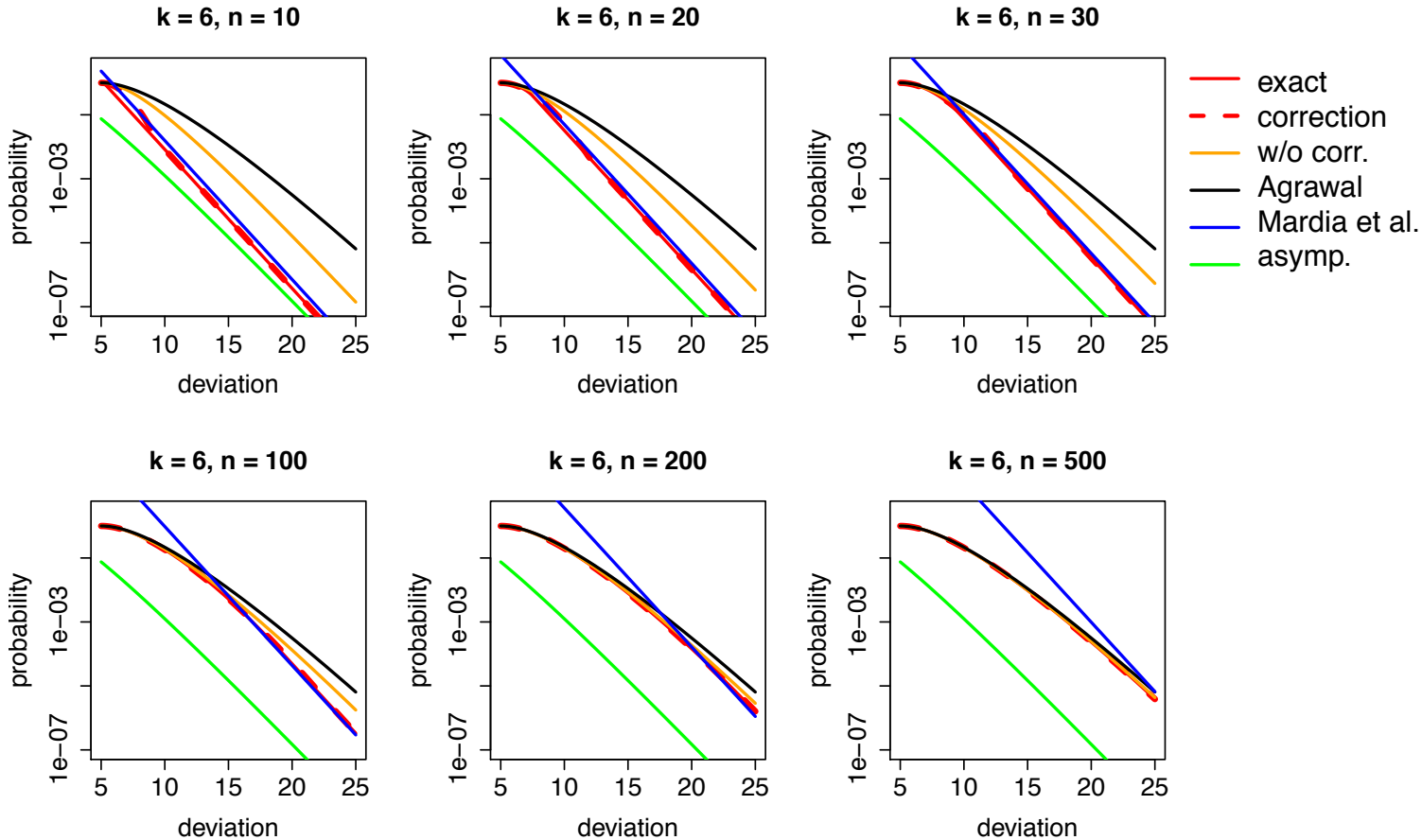
$$\mathbb{P}\left(n\,\mathcal{D}(\hat{p}_{k,n}\|p) > t\right) \leqslant e^{-t}e^{k-1}G_{k,n}\left(1-\frac{k-1}{t}\right). \quad (11)$$

- (Better idea) One step Newton iteration:
  For $n \geqslant 1$, $k \geqslant 2$ and $t > k-1$,

$$\mathbb{P}\left(n\,\mathcal{D}(\hat{p}_{k,n}\|p) > t\right) \leqslant \exp(-\hat{\lambda}_{k,n}t)G_{k,n}(\hat{\lambda}_{k,n}). \quad (12)$$

  where $\hat{\lambda}_{k,n} := \min\left\{1-\frac{k-1}{t}+\frac{k}{k-1}\frac{t-k+1}{n}, 1\right\}$.

New bounds are competitive with the state-of-the art



Bounds on $\mathbb{P}(n\,\mathcal{D}(\hat{p}_{k,n}\|p) > t)$ for $k = 6$ and $t > \min(\log G_{k,n}(1), k-1)$. The y-axis is in logarithmic scale.

The methods compared include: "exact" (from numerical minimization), "correction", "w/o correction", Agrawal (2020), Mardia *et al.* (2019), and the asymptotic bound that is the exact probability when $n \to \infty$; the latter will not be a valid bound in general and is for reference only.

Back to the Lipid Data . . .

For the lipid data this allows us to obtain a 95% interval for the *ignorance region*: $[0.151, 0.907]$
In other words, with coverage probability 95% it will contain the interval $[\mathrm{L}, \mathrm{U}]$, where L (U) is the lower (upper) bound on the ACE compatible with the population distribution of the observables.

In constructing this interval we have also directly tested the IV restrictions in the following sense:
Under the IV model,

$$\mathrm{P}(\text{ model not rejected, and interval contains the ACE}) \geqslant 95\%$$

Falsification test is built in: *If the interval is empty then the IV model is rejected!*

Summary

- Instrumental Variable (IV) Models represent an approach to causal analysis of imperfect experiments
- Derived bounds for the IV model by viewing the model as two observational studies to which participants are randomly assigned;
- Presented a Bayesian approach via a transparent parametrization that separates identified and non-identified parameters.

Some References

- M. Chang, S. Lee, Y.-J. Whang (2015). Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. Econometric Journal
- M. Gail, R. Simon (1985). Testing for qualitative interactions between treatment effects and patient subsets. Biometrics.
- F.R. Guo, T.S. Richardson (2021). Chernoff-type concentration of empirical probabilities in relative entropy. *IEEE Transactions on Information Theory*.
- D. Kédagni, I. Mourifié. (2019). Generalized instrumental inequalities: Testing the IV independence assumption. *Biometrika*.
- L. Wang, J.M. Robins, T.S Richardson. (2017). On Falsification of the Binary Instrumental Variable Model. *Biometrika*.
- T.S. Richardson, R.J. Evans, J.M. Robins. (2011). Transparent Parameterizations of Potential Outcome Models (with discussion). *Bayesian Statistics 9*