COLUMBIA UNIVERSITY
Department of Biomedical Informatics

The Counterfactual χ-GAN

Amelia J. Averitt[1] Natnicha Vanitchanant[1] Rajesh Ranganath[2] Adler J. Perotte[1]

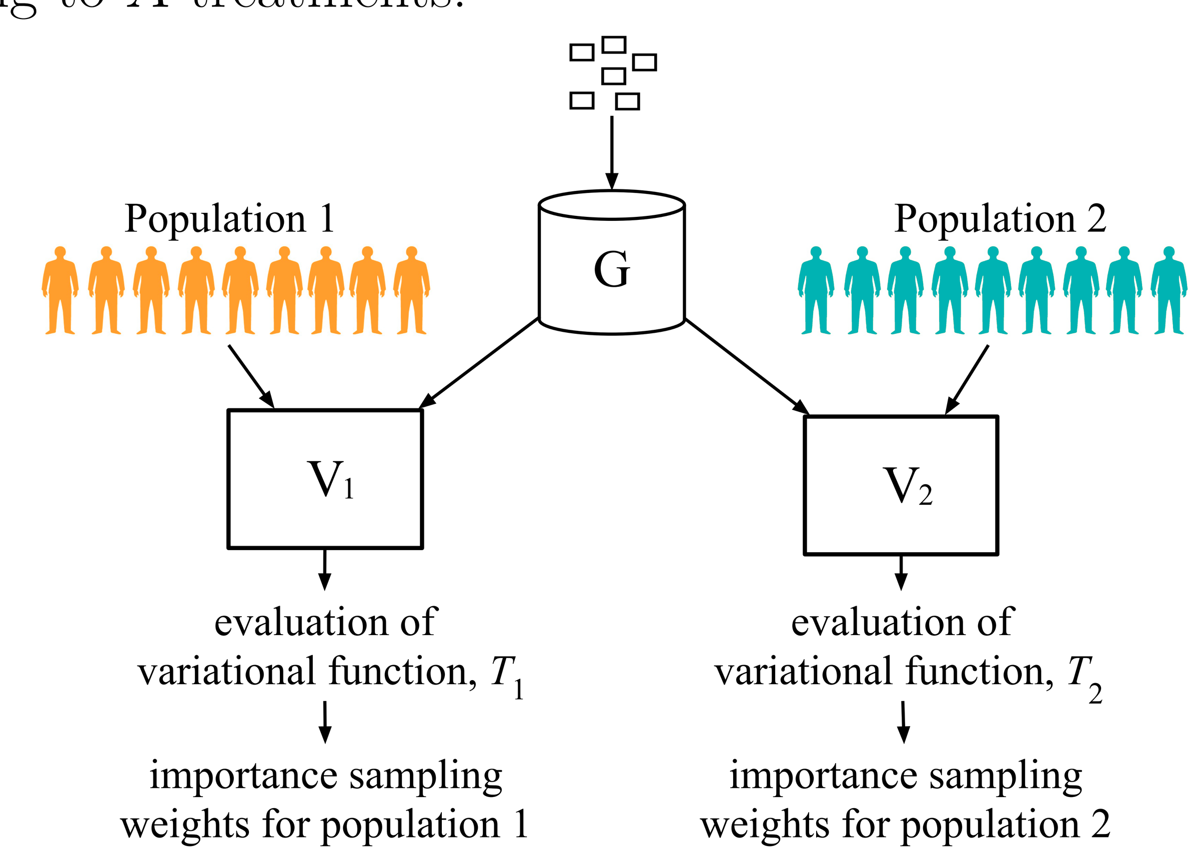
[1]Columbia University, Department of Biomedical Informatics [2]New York University, Courant Institute of Mathematical Sciences

Introduction

Counterfactual causal inference requires the assumption of strong ignorability. Observational data typically does not satisfy this assumption. Weighting is a method to enforce strong ignorability, but many weighting methods estimates can be unstable. [1]

The Model. This research proposes a generative adversarial network (GAN)-based model called the Counterfactual χ-GAN (cGAN) to learn stable, feature-balancing weights. [2]

- cGAN utilizes the f-GAN framework to minimize the χ-divergence. [3]
- there is a single generator, G, and two or more variational functions V_a, corresponding to A treatments.



N_a units are drawn from an unknown and population-specific distribution q_a. Our objective is to identify a set of weights, w_a, for each population that allow estimation of expectations from the same target distribution, p.

$$\sum_{a=1}^{A} \chi^2 \left(p(x) \parallel q_a(x)\right)$$

Of Note About the Model

- cGAN uses the variational divergence minimization to optimize the χ-divergence
- G identifies the target distribution that minimizes the sampling variance and encourages coverage
- Weights (w_a) are directly calculated from the output of V_a and enforce the assumption of strong ignorability.

Experimentation

Simulation. To evaluate cGAN, we applied the model to a simulation in which the ground truth is known. Two populations were constructed such that half of each was generated from the same distribution (Subpop A).

- Pop 1 = Subpop A + Subpop B
- Pop 2 = Subpop A + Subpop C

cGAN learned higher weights for units from Subpop A than units not generated from this subpopulation.

Application to Clinical Data. We additionally applied cGAN to a clinical experiment using EHR data from Columbia University Irving Medical Center.

- sitagliptin vs. glimepiride in elderly patients with T2DM

We evaluated cGAN in improving feature balance compared to inverse propensity weighting (IPW), clipped-IPW, using the absolute standardized difference in the means (ASDM)

Method	ASDM
unweighted	0.1103
IPW	0.0876
clipped-IPW	0.0631
cGAN	**0.0252**

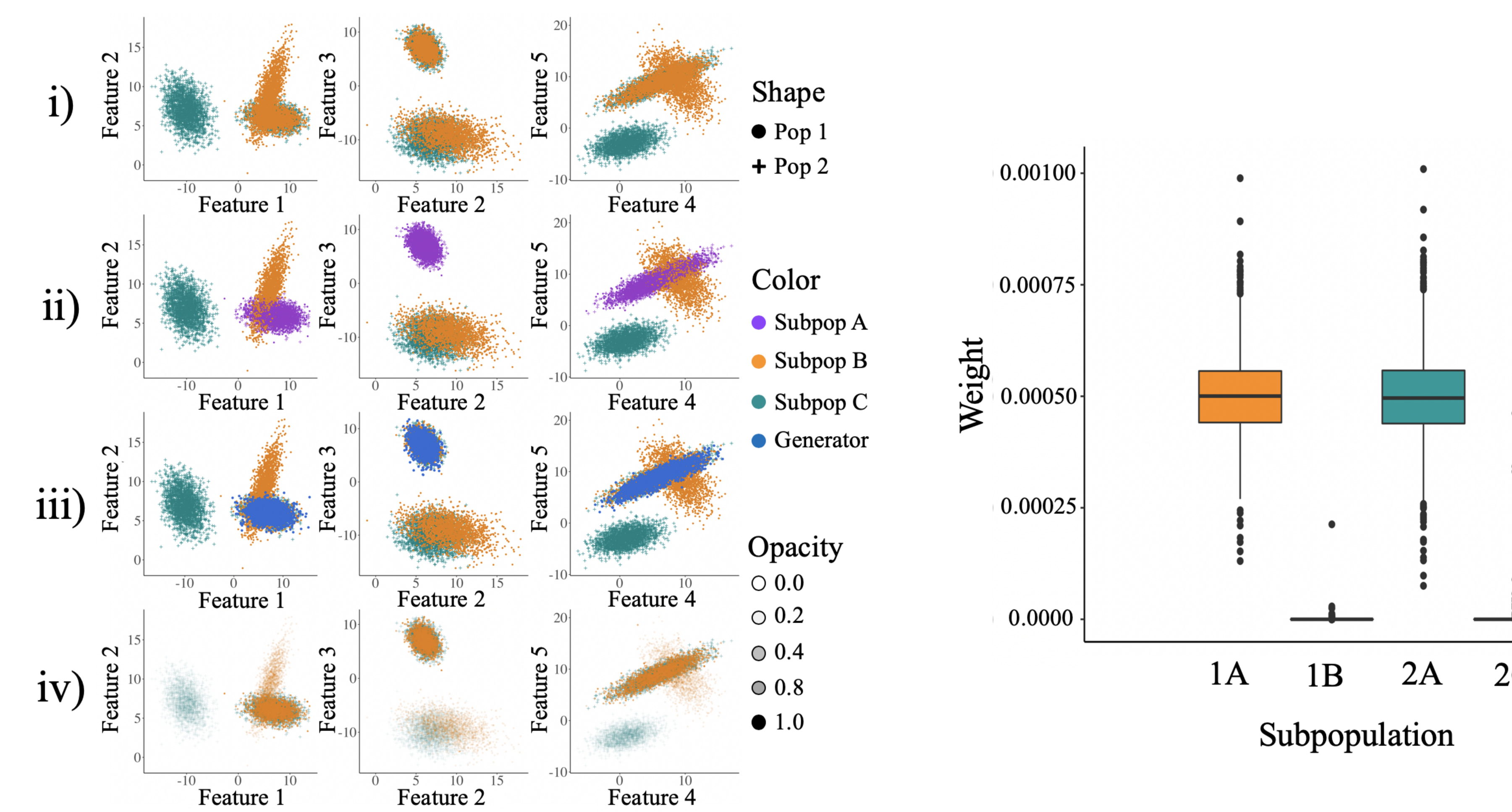


Figure 1: Simulation Results. Left: Select features (i) by population of origin; (ii) with subpopulation A highlighted; (iii) samples from the generator; (iv) opacity adjusted by weight. Right: Distribution of weights by subpopulation.

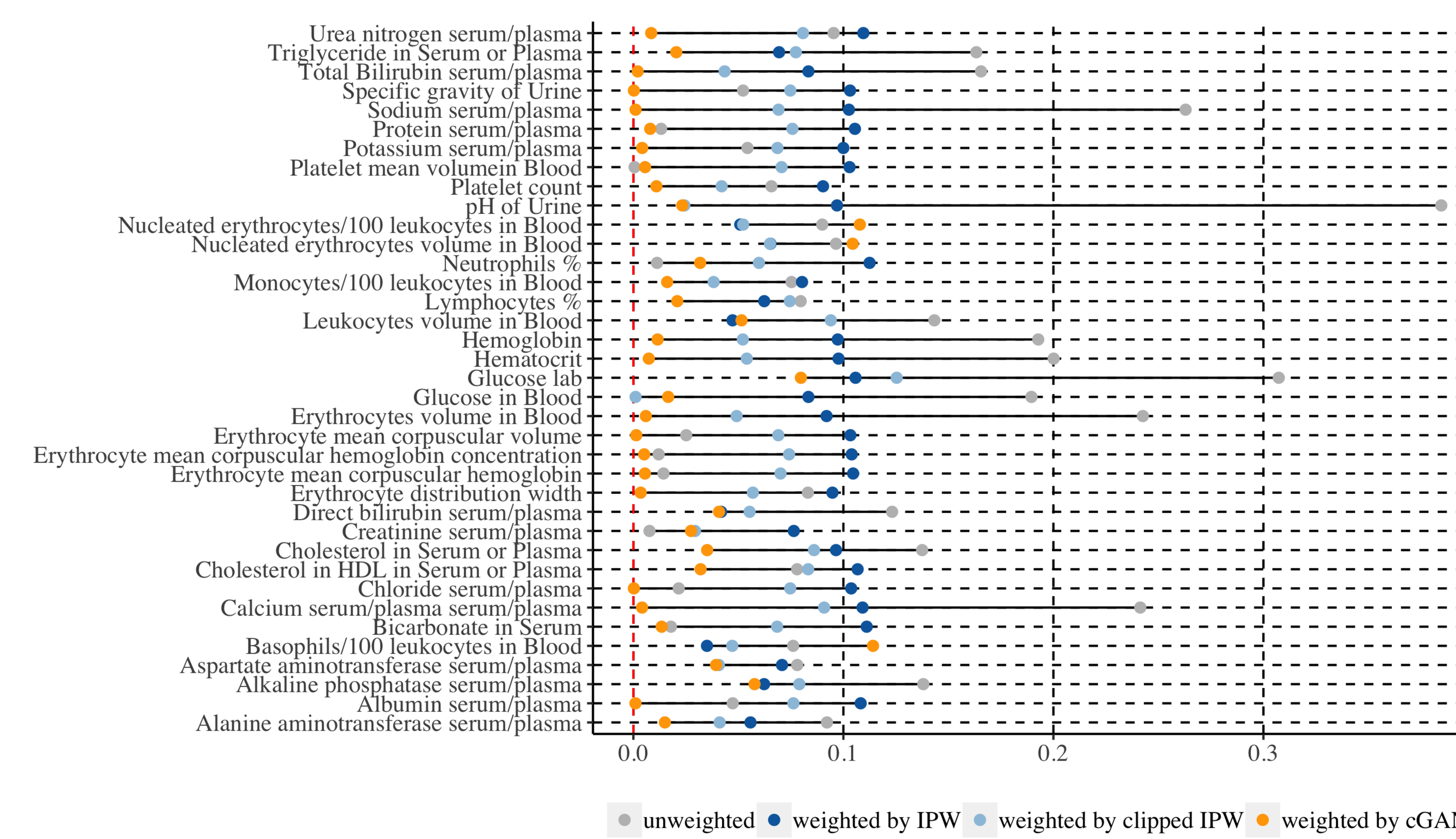


Figure 2: ASDM for Clinical Data

[1] Rubin, D. B. *Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688–701 (1974).*

[2] Goodfellow, I. J. *Generative Adversarial Networks. (2014).*

[3] Nowozin, S. Cseke, B. Tomioka, R. *f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. NIPS. (2016).*