Comments on Simulation Programs

Here I am following the notation in the matlab code, with parenthetical comments about the corresponding concepts in the paper.

For each of the designs there are n units in the population. Each unit i is characterized by three variables, Z_i, α_i, and ϕ_i (the $\theta_{n,i}$ in the paper - the notation in the programs should go back to this). The potential outcome function is

$$Y_i(x) = \alpha_i + \phi_i x.$$

(It would have been more proper to write it as a function of u, and work from there.)

Z_i is a $KZ+1$ dimensional vector with the first element equal to 1, and the remaining KZ elements drawn from a normal distribution with mean zero and unit variance, all independent. The population is generated first, with

$$Z_{i1} = 1, \qquad Z_{ik} \sim \mathcal{N}(0,1).$$

$$\alpha_i \sim \mathcal{N}(Z_i^\top \gamma, \sigma_\eta^2)$$

$$\phi_i \sim \mathcal{N}(Z_i^\top \psi, \sigma_\varepsilon^2).$$

The assignment mechanism is

$$X_i \sim \mathcal{N}(Z_i^\top \xi, 1).$$

In all designs $\xi = 0$, so that for the properties the U_i and X_i are interchangeable.

Once we have this population, we can calculate the various estimands and the various variances. Because X_i and Z_i are uncorrelated, for the limit of the estimands we have

$$\lim_{n\to\infty} \theta_n = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \phi_i = \mathbb{E}[Z]^\top \psi = \psi_1 = 0$$

in all the designs.

$$\lim_{n\to\infty} \gamma_n = \mathbb{E}[Z_i Z_i^\top]^{-1} \mathbb{E}[Z_i Y_i] = \mathbb{E}[Z_i Y_i].$$

For the first element of the limit of γ_n, corresponding to the intercept, we have

$$\lim_{n\to\infty} \gamma_{1n} = \mathbb{E}[Y_i] = \mathbb{E}[\alpha_i] + \mathbb{E}[\phi_i X_i] = 0 + \mathbb{E}[Z_i^\top \psi_i X_i] + \mathbb{E}[(\phi_i - Z_i^\top \psi) X_i] = 0$$

For the other elements

$$\lim_{n\to\infty} \gamma_{kn} = \mathbb{E}[Z_{ik} Y_i] = \mathbb{E}[Z_{ik}\alpha_i] + \mathbb{E}[Z_{ik}\phi_i X_i]$$

$$= \mathbb{E}[Z_{ik}Z_i^\top \gamma] + \mathbb{E}[Z_{ik}(\alpha_i - Z_i^\top \gamma)] + \mathbb{E}[Z_{ik}Z_i^\top \psi_i X_i] + \mathbb{E}[Z_{ik}(\phi_i - Z_i^\top \psi)X_i] = 0$$

==

Next, we simulate samples. This involves samping units from the population, and assigning them causes X_i to the units sampled.

The list of programs used in the simulations is

1. `main_18aug8.m` This is the main program that runs the simulations.
2. `gen_potential.m` This program generates the population. It takes as input the population size n, the $KZ+1$ vector ψ, the $KZ+1$ vector γ, the $KZ+1$ vector ξ, and the scalars σ_ε, and σ_η, and the integer KZ, and puts out the $KZ+1$ vector Z_i, and the scalars α_i, and ϕ_i, for $i = 1, \ldots, N$.
    (a) KZ is the number of characteristics, beyond the intercept. In the first design $KZ = 1$.
    (b) ξ is not used
    (c) In the first design, $\gamma = (0, 0)^\top$.
    (d) In the first design, $\psi = (0, 2)^\top$.
    (e) $\alpha_i \sim \mathcal{N}(Z_i^\top \gamma, \sigma_\eta^2)$
    (f) $\phi_i \sim \mathcal{N}(Z_i^\top \psi, \sigma_\varepsilon^2)$
3. `gen_sample.m` This program takes the population, characterized by (Z_i, α_i, ϕ_i, and samples a fraction ρ from this population, and assigns a value of X_i to them. The input ξ ??? The output is
    (a) YR
    (b) XR
    (c) ZR
    (d) W
    (e) UR
    (f) R
    (g) Ω
    (h) θ_{desc}
    (i) $\theta_{\text{causal,sample}}$