Assumptions of the SDir Method

2025-04-22

Formalizing SD_{ir}

Let x_i be 0 when subject i receives the control intervention or 1 when they receive the experimental intervention. α_i is the effect of the control and β_i is the effect of the experimental intervention for subject i. In a trial, we are interested in estimating the experimental intervention's effect relative to the control's effect ($\gamma_i = \beta_i - \alpha_i$):

$$\begin{aligned} \text{post}_i &= \text{pre}_i + \alpha_i\,(1 - x_i) + \beta_i\,x_i \\ (\text{post}_i - \text{pre}_i) &= \alpha_i + (\beta_i - \alpha_i)\,x_i \\ \delta_i &= \alpha_i + \gamma_i\,x_i \end{aligned}$$

Nota bene the heterogeneity of treatment effects is $\text{Var}[\gamma]$.

$$\begin{aligned} \text{Var}[\delta \mid x = 0] &= \sigma_\alpha^2 \\ \text{Var}[\delta \mid x = 1] &= \sigma_\alpha^2 + \sigma_\gamma^2 + 2\,\rho_{\alpha,\gamma}\,\sigma_\alpha\,\sigma_\gamma \end{aligned}$$

The difference between these variances is $\text{Var}[\delta \mid x = 1] - \text{Var}[\delta \mid x = 0] = \sigma_\gamma^2 + 2\,\rho_{\alpha,\gamma}\,\sigma_\alpha\,\sigma_\gamma$. This implies that $\sigma_\beta^2 = \sigma_\alpha^2 + \sigma_\gamma^2$ when $\rho_{\alpha,\gamma} = 0$, which is SD_{ir}^2.

What does $\rho_{\alpha,\gamma} = 0$ imply about the correlation between counterfactual outcomes (α and β)?

$$\begin{aligned} \gamma_i &= \beta_i - \alpha_i \\ \implies \sigma_\gamma^2 &= \sigma_\beta^2 + \sigma_\alpha^2 - 2\,\rho_{\beta,\alpha}\,\sigma_\beta\,\sigma_\alpha \\ \implies \sigma_\beta^2 - \sigma_\alpha^2 &= \sigma_\beta^2 + \sigma_\alpha^2 - 2\,\rho_{\beta,\alpha}\,\sigma_\beta\,\sigma_\alpha \\ \implies \rho_{\beta,\alpha} &= \frac{\sigma_\alpha}{\sigma_\beta} \end{aligned}$$

This is a strong assumption. First, it implies that negative correlations between potential outcomes are impossible, but there are reasonable cases when they can be expected (see below). Second, it implies that the treatment effect (γ) is orthogonal to the control effect (α), which may be a reasonable assumption in some but certainly not all experiments. Here, our goal is to dissect this assumption to provide a couple of examples of when it may and may not hold.

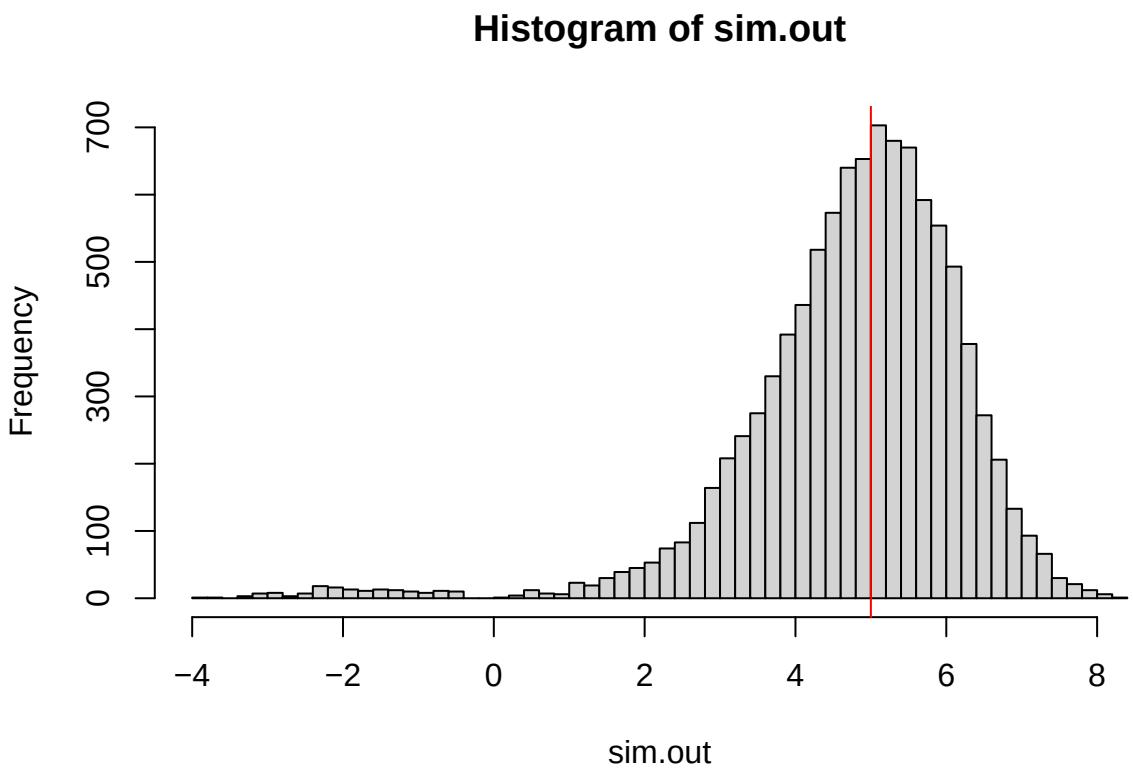
When are SD_{ir}'s assumptions reasonable?

One may be reasonably comfortable with the assumptions made by the SD_{ir} method and its associated data-generating model in a few different experimental contexts. In their original papers, XXXXX described parallel group experiments in which no or negligible changes were expected in the control group, while change was expected in the intervention group. Such contexts are arguably cases when SD_{ir}'s assumptions would be reasonable. For instance, in exercise science, a common relevant study would be one where inactive, healthy

adults are randomized to one of two groups: non-exercise control or a resistance training intervention, which aims to answer the question: How does resistance training in healthy, inactive adults affect isometric knee extension strength? Suppose that these inactive adults are in a relatively stable period where their strength levels can be assumed to be stationary on average, albeit with some random fluctuations. The investigator may be fine with the principal assumption of the SD_{ir} method: Effects of the intervention are independent of *changes* that would occur in the control condition.

We can simulate such a study and show that the SD_{ir} provides (relatively) unbiased estimates of treatment effect heterogeneity (NB, a bias exists insofar as standard deviation is biased due to Jensen's inequality).

```
nsim <- 1e4
sigma_meas <- 5
n <- 100
D <- 20
sdir <- 5

calc.sdir <- function(x,y) sign(var(y)-var(x))*sqrt(abs(var(y)-var(x)))

sim.out <- c()
for(i in 1:nsim) {
  control_deltas <- rnorm(n,0,5)
  exp_deltas <- control_deltas + rnorm(n,D,sdir)

  obs_control <- control_deltas + rnorm(n,0,sigma_meas)
  obs_exp <- exp_deltas + rnorm(n,0,sigma_meas)

  sim.out <- c(sim.out,
               calc.sdir(obs_control, obs_exp))
}

hist(sim.out,breaks="fd")
abline(v=sdir, col="red")
```

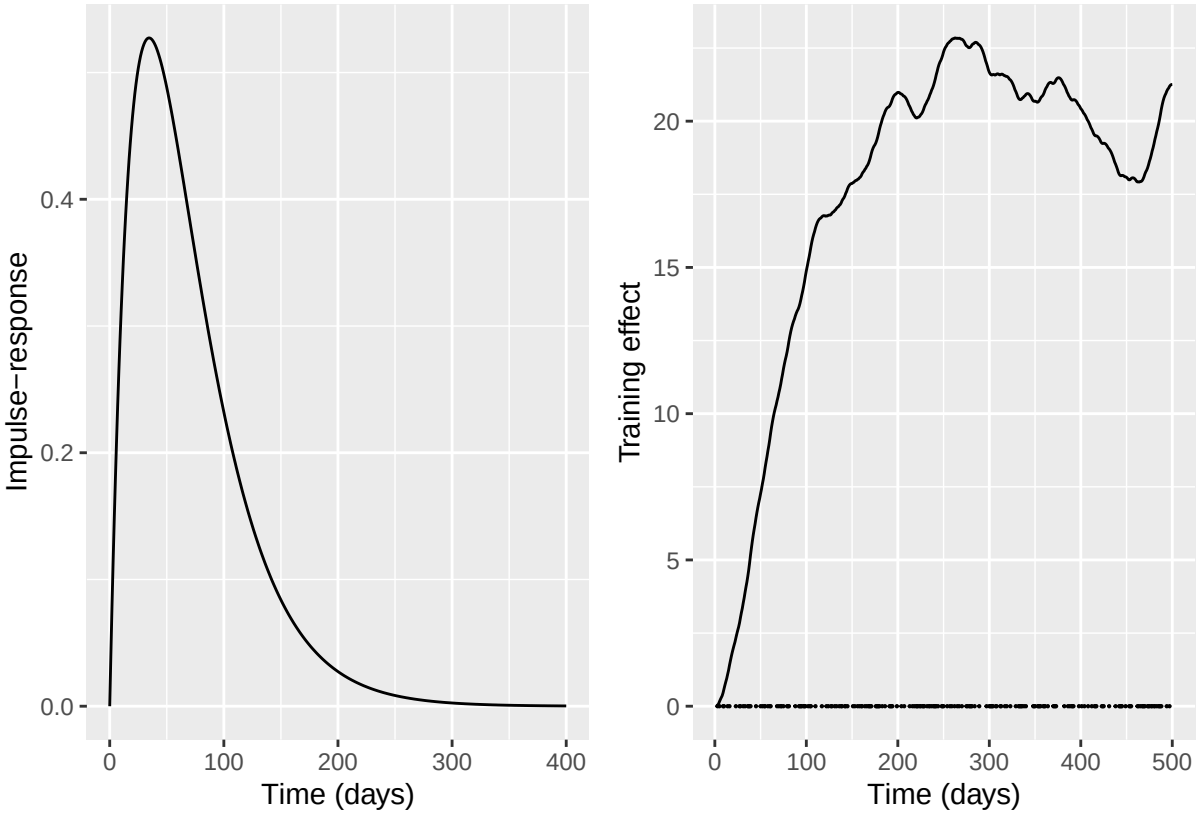



```
print(mean(sim.out))
```

```
## [1] 4.786248
```

When are SD_{ir}'s assumptions unreasonable?

We'll now turn to an example in which the SDir assumptions do not hold. Suppose adaptations to a single training session are defined by the impulse-response function (left), and repeated exercise sessions result in cumulative adaptation (right).



We would like to study the effects of exercising approximately three times per week relative to no exercise. We will recruit from the general population, meaning that the individuals recruited will have varying levels of training experience. This training experience is defined by the frequency of training, in which participant i's probability of training on a given day is $p_i \sim \text{Beta}(0.4, 0.8)$; the cumulative training effect from the 500 days before enrollment was then simulated and used as the pre-intervention score.

When participant i is randomized, they will either exercise with a probability of $\frac{3}{7}$ (intervention) or 0 (control) for 60 days. Thus, those who were well-trained prior to enrollment will experience little-to-no change (intervention) or detraining effects (control). Conversely, those who were untrained prior to enrollment will experience training effects (intervention) or little-to-no change (control).

```
sims <- pbmcapply::pbmclapply(1:50, function(x) {
  nsub <- 500
  probs <- rbeta(nsub, 0.4, 0.8)
  df <- c()
  for(i in 1:nsub) {
    trained <- rbinom(500,1,probs[i])

    trained0 <- c(trained,rbinom(60,1,0))
    trained1 <- c(trained,rbinom(60,1,3/7))

    ts <- 1:length(trained0)

    convolved0 <- pracma::conv(
      irf(ts, 5, 30, 40),
      trained0
    )[1:length(ts)]
```

```
    convolved1 <- pracma::conv(
      irf(ts, 5, 30, 40),
      trained1
    )[1:length(ts)]

    N <- length(ts)

    df <- rbind(df, data.frame(id = i,
                               group = as.numeric(i > nsub/2),
                               pre = convolved0[500],
                               post0 = convolved0[N],
                               post1 = convolved1[N]))
  }

  df$post <- ifelse(df$group,df$post1,df$post0)
  df$delta <- df$post - df$pre

  control <- subset(df, group == 0)
  exp <- subset(df, group == 1)

  plot(df$post0-df$pre, df$post1-df$pre)

  #summary(lm(post ~ pre + group, df))

  # double-check this output!
  data.frame(
    var_ir = var(exp$delta) - var(control$delta),
    varD = var( (df$post1-df$pre) - (df$post0-df$pre) ),
    rho = cor( df$post0-df$pre , df$post1-df$pre ), # correlation between deltas
    rho1 = cor( (df$post0-df$pre) , df$post1-(df$post0-df$pre) ) # correlation between control delta an
    )
}, mc.cores = 8)

colMeans( do.call(rbind, sims) )
```

```
##     var_ir       varD        rho       rho1
##  3.6615489  2.9479240  0.9793796 -0.9929880
```