RISE: Randomization Inference for Spillover Effects

Mark M. Fredrickson

University of Illinois, Urbana-Champaign

http://www.markmfredrickson.com

Randomization Inference

Rather than assume a distributional form on the outcome R, randomization inference takes the assignment to treatment Z as the primary random quantity understudy. Given two potential outcomes Y_t and Y_c, $R = ZY_t + (1 - Z)Y_c$. Inference is performed using all possible treatment assignments Ω (Rosenbaum, 2010, ch. 2). We can also consider an algorithmic interpretation (Bowers, 2011):

- The **model of treatment assignment** defines the universe of possible treatment assignments (i.e. how Z takes values).
- The **model of effects** is a function that maps an outcome and a treatment assignment to the set of adjusted outcomes. The *additive model of effects* states that $Y_t = Y_c + \tau$. A function that generates additive models of effect given τ:

```
additive.effect <- function(tau) {
    function(responses, z) {
        responses - z * tau
    }
}
```

- The **test statistic** is a function that maps a set of outcomes and a treatment assignment to a real number. The test statistic summarizes the differences between treatment and control groups. The difference of means test statistic:

```
difference.of.means <- function(ys, z) {
    mean(ys * z) - mean(ys * (1 - z))
}
```

Hypothesis test for a parameter of the model of effects:

1. Let Ω be all possible randomizations as defined by the model of treatment assignments. Let D be a vector of size $|\Omega|$
2. For each z in Ω
   (a) Adjust the observed outcomes R using the model of effects and z.
   (b) Compute the test statistic on the adjusted outcomes and store in D.
3. Compute the test statistic t on the observed outcomes.
4. The p-value of the parameter is: $2\min\left(\frac{|\{d\in D: d\geq t\}|}{|D|}, \frac{|\{d\in D: d\leq t\}|}{|D|}\right)$.

Testing models of spillover

Models of effects may rely on any covariate, for example a fixed network S between units. We can write a model of effects that provides a constant change in the outcomes for each treated unit in a neighborhood: $Y_t = Y_c + \sigma Z^t S$. In this equation there are $|\Omega|$ possible potential outcomes, each corresponding a randomization (Rosenbaum, 2007). A function to create such models of effect for a given σ as:

```
network.effect <- function(sigma) {
    function(responses, z) {
        responses - z %*% S * sigma
    }
}
```

We can also compose models of effect. For example, we can combine the additive and network models of effect as $Y_t = Y_c + \tau + \sigma Z^t S$. A generator function for this model:

```
combined.effects <- function(tau, sigma) {
    f1 <- additive.effect(tau)
    f2 <- network.effect(sigma)
    function(responses, z) {
        f1(f2(responses, z), z)
    }
}
```

Simulated Experiment

Using the model of effects above, we can simulate an experiment in which spillover effects play a role. Let our subject pool be $n = 40$ units, with 20 units assigned to treatment. The covariates are fixed quantities that are randomly generated for convenience.

```
covariates <- function(n = 100) {
    x1 <- rnorm(n, 0, 10)
    x2 <- rbeta(n, 7, 1)
    data.frame(x1, x2)
}
n <- 40
X <- covariates(n)
Z <- randomize(n)
```

The network S is fixed, though again generated through a random process for convenience. The network is an adjacency matrix where any edge exists with $p = 0.04$.

```
S <- network(n, 0.04)
```

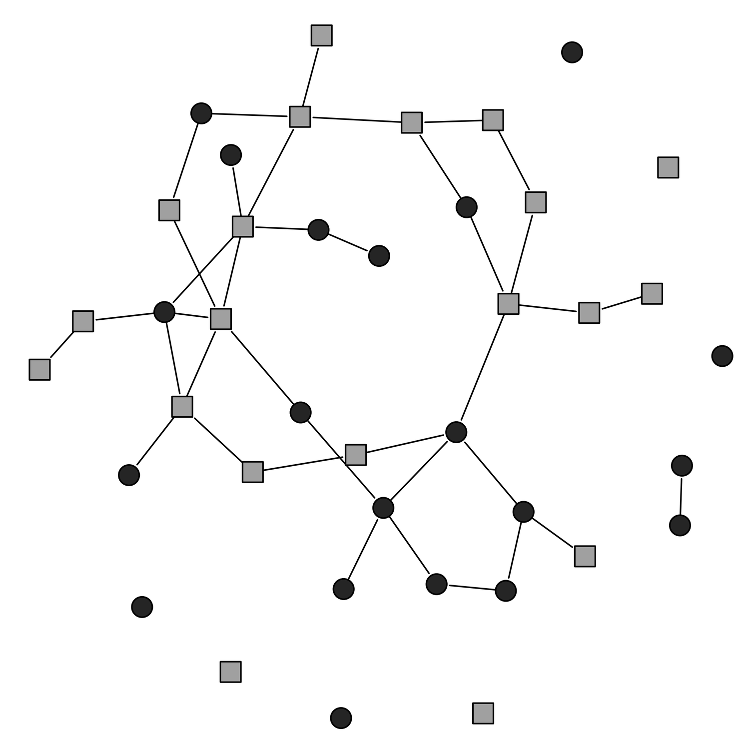

Figure 1: *Network: treatment (circles) and control (squares)*

I create Y_t from the covariates and then adjust it using the model of effects to recover Y_c. I choose true values for the parameters: $\tau = 5, \sigma = 1$.

```
outcomes <- function(covars, tau, sigma, z) {
    yt <- with(covars, x1 * x2)
    yc <- yt - tau - sigma * as.vector(z %*% S)
    data.frame(yc, yt)
}
Y <- outcomes(X, 5, 1, Z)
R <- ifelse(Z, Y$yt, Y$yc)
```

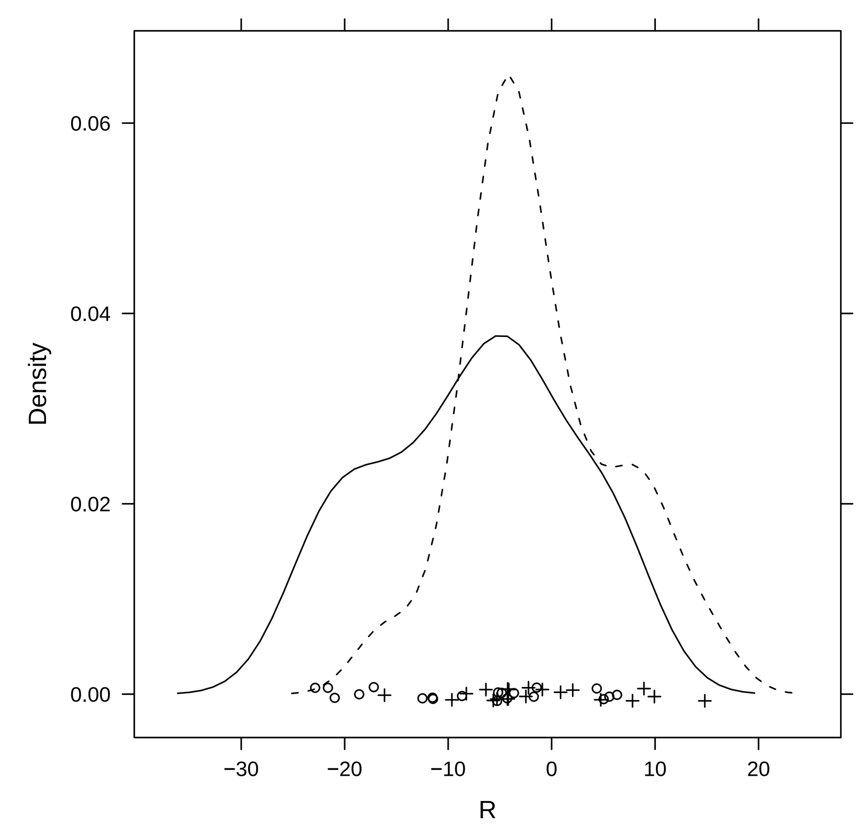

Figure 2: *Observed outcome by treatment (dotted) and control (solid)*

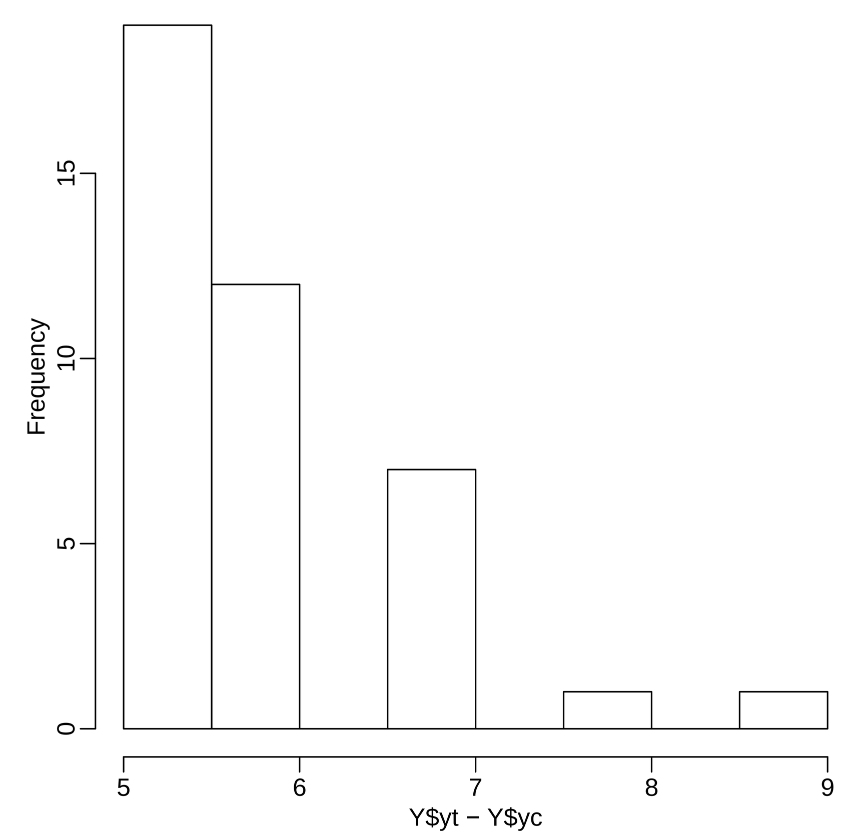

Figure 3: *Distribution of $Y_t - Y_c$*

With 40 subjects of which 20 are treated there are 137846528820 possible randomizations, so we will sample 1000. For each sample, we evaluate a set of plausible hypotheses for τ and σ.

```
results <- parameterizedRandomizationDistribution(
  R, Z,
  difference.of.means,
  combined.effects,
  list(tau = seq(-20, 15, 0.5),
       sigma = seq(-10,10,0.5)),
  samples = 1000)
results.ci <- confint(results, level = 0.95)
```

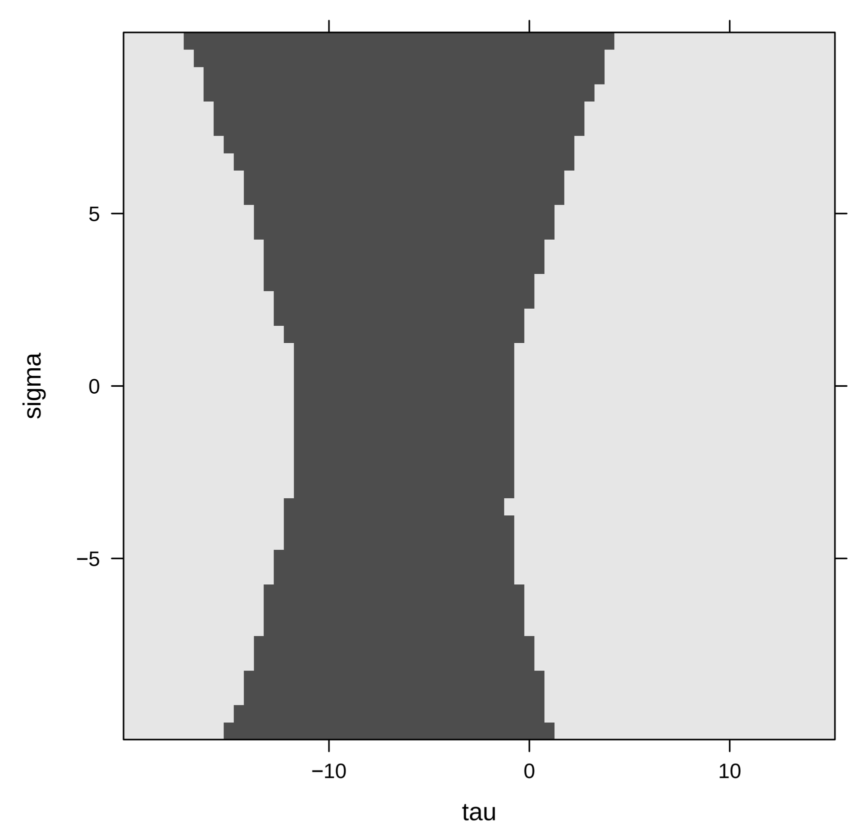

Figure 4: *Joint confidence region of τ and σ.*

Atlantic Causal Inference Conference, 2011, University of Michigan

References

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191 – 200.

Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer, New York.