Estimating Spatial Effects

Cyrus Samii (NYU),
joint with Peter M. Aronow (Yale)

March 5, 2018

Motivation

- Political campaign deploys events to certain places. Content spreads via word of mouth.
- Public health intervention applies treatments at certain locations. “Herd immunity” dissipates in distance from these locations.
- Forest rangers establish monitoring stations on the forest edge. Deterrence effects dissipate in distance from stations.
- Child registration campaign up sets registration sites in schools. Effect on household registration depends on distance from site.

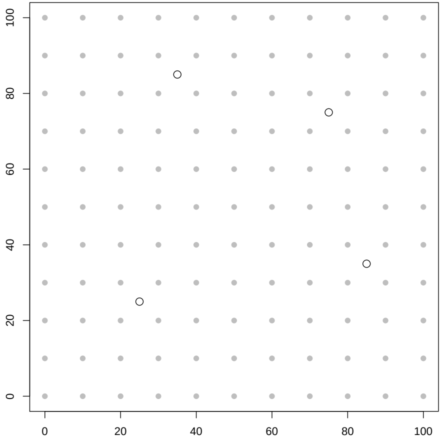
Motivation

- Political campaign deploys events to certain places. Content spreads via word of mouth.
- Public health intervention applies treatments at certain locations. "Herd immunity" dissipates in distance from these locations.
- Forest rangers establish monitoring stations on the forest edge. Deterrence effects dissipate in distance from stations.
- Child registration campaign up sets registration sites in schools. Effect on household registration depends on distance from site.
- Apply a *design-based* perspective.
  - General with respect to outcome DGP.
  - Analysis focuses on what we can *control* (design).

Motivation

- A common assumption in causal inference is "no interference":
  - Part of the "stable unit treatment value assumption" (SUTVA)
  - For treatment vector $\mathbf{z}$, potential outcomes are as $Y_i(\mathbf{z}) = Y_i(z_i)$.
- May not make sense in the applications described above.

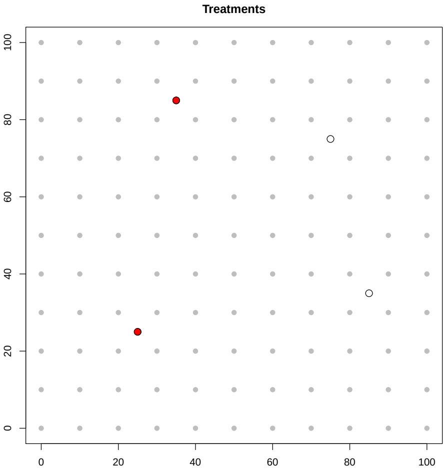
Setting



- Set $\mathcal{N}$ of intervention nodes indexed by $i = 1, \ldots, N$.
- On a field $\mathcal{X}$ of points indexed by $x = (x_1, x_2)$.
- Binary treatments assigned to intervention nodes, $\mathbf{Z}$.
- Potential outcomes at x for each value of $\mathbf{Z}$, $Y_x(\mathbf{z})$.
- Observed outcomes at x as a function of potential outcomes:

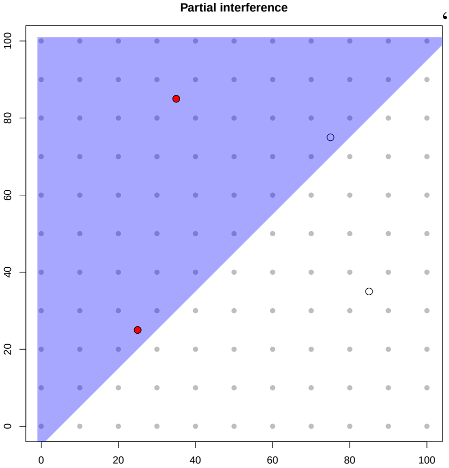
$$Y_x = \sum_{\mathbf{z} \in \mathcal{Z}} Y_x(\mathbf{z}) I(\mathbf{Z} = \mathbf{z})$$

Setting



- We design an experiment and assign $\mathbf{Z}$ at random.
- Observed outcomes are then, $Y_x(\mathbf{Z})$ for $x \in \mathscr{X}$.
- We suspect interference.
- What causal leverage does a randomized design nonetheless yield?

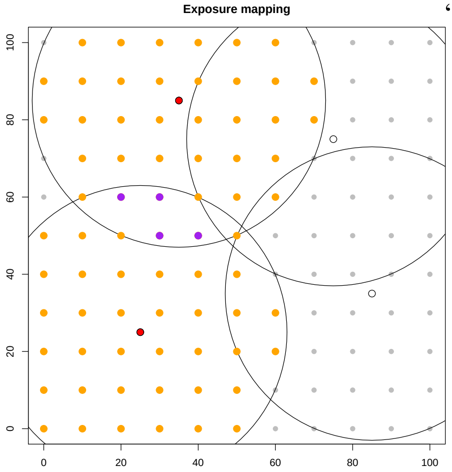
Existing approaches



"Partial interference" approach:

- Hudgens & Halloran (2008)
- Arbitrary interference within strata, but none between.
- Two-stage randomization design:
  - Saturation between strata
  - Treatment location within
- Design yields consistent inference about effects of more or less "direct" exposure.

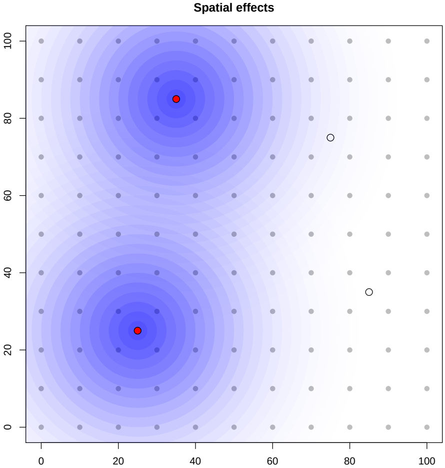
Existing approaches



"Exposure mapping" approach:

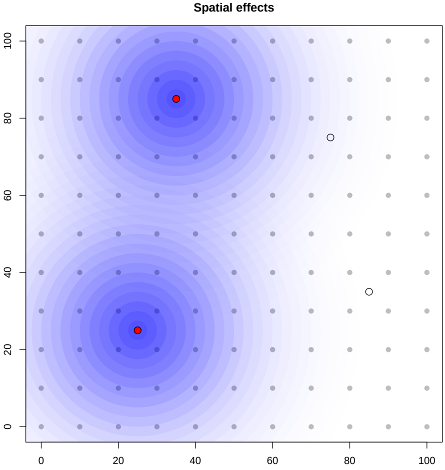
- Aronow & Samii (2017)
- Suppose $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$ for classes of $(\mathbf{z}, \mathbf{z}')$ (level sets).
- Variations for which $Y_i(\mathbf{z}) \neq Y_i(\mathbf{z}')$ for $\mathbf{z} \neq \mathbf{z}'$ are known exactly—"exposures."
- Design yields consistent inference about average exposure effects.
- Adjust for design-induced probabilities of exposure.

Problems with current approaches



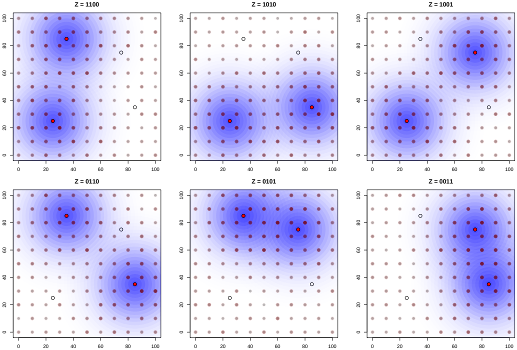
- Spillovers may bleed out in ways that violate partial interference.
- Exposures may vary over all treatment profiles (no true level sets).

Strategy



- Shift the goal posts: define an effect that *is* identified by the randomized design in this setting.
- Study this effect to see how it may align with effects of substantive interest.
- Draw out implications for design so as to bring about such an alignment.

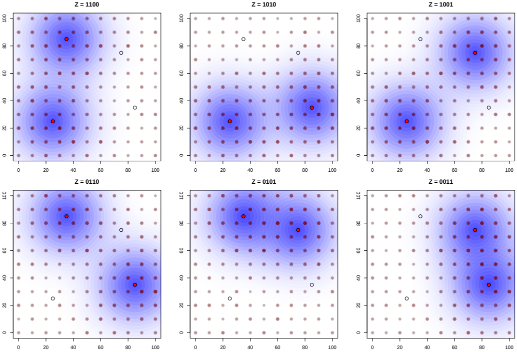
Setting



The design is as follows:

- Random assignment: $\mathbf{Z} \perp\!\!\!\perp \mathscr{G}\{(Y_x(\mathbf{z}), W_x) : x \in \mathscr{X}, \mathbf{z} \in \mathscr{Z}\}$, where $\mathscr{G}$ is an arbitrary mapping to $\mathbf{Z}$'s probability space.
- Positivity: $0 < \Pr(\mathbf{Z} = \mathbf{z} | \alpha) < 1$ for all $\mathbf{z} \in \mathscr{Z}$.
- Observe $\mathbf{Z}$ and $(\mathbf{Y}_{\mathscr{X}}, \mathbf{W}_{\mathscr{X}})$.

Setting

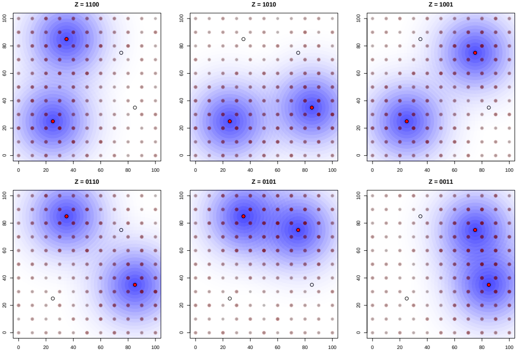


- A useful decomposition:

$$
\begin{aligned}
Y_x &= \sum_{\mathbf{z} \in \mathscr{Z}} Y_x(\mathbf{z}) I(\mathbf{Z} = \mathbf{z}) \\
&= Z_i \sum_{\mathbf{z}_{-i} \in \mathscr{Z}_{-i}} Y_x(1, \mathbf{z}_{-i}) I(\mathbf{Z}_{-i} = \mathbf{z}_{-i}) + (1 - Z_i) \sum_{\mathbf{z}_{-i} \in \mathscr{Z}_{-i}} Y_x(0, \mathbf{z}_{-i}) I(\mathbf{Z}_{-i} = \mathbf{z}_{-i})
\end{aligned}
$$

- Outcome at x when node i is switched on, and when node i is switched off.

Defining effects

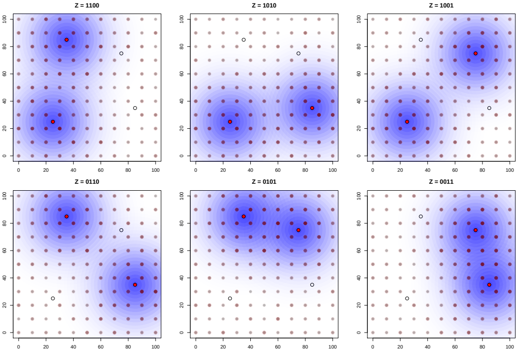


- "Individualistic" average of potential outcomes for point x, holding node i to treatment value z:

$$Y_{ix}(z;\alpha) = \sum_{\mathbf{z}_{-i} \in \mathscr{Z}_{-i}} Y_x(z, \mathbf{z}_{-i}) \Pr(\mathbf{Z}_{-i} = \mathbf{z}_{-i} | Z_i = z, \alpha)$$

  with α governing distribution of $\mathbf{Z}$.
- E.g., mean of top row for $Z_1 = 1$.

Defining effects

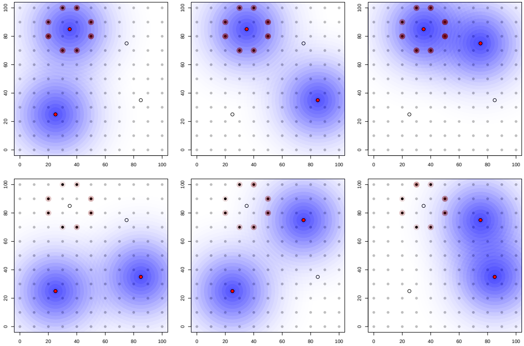


- Point-specific effect at x of switching treatment at node i:

$$\tau_{ix}(\alpha) = Y_{ix}(1;\alpha) - Y_{ix}(0;\alpha).$$

- E.g., diff. in means between top and bottom rows for Z_1.

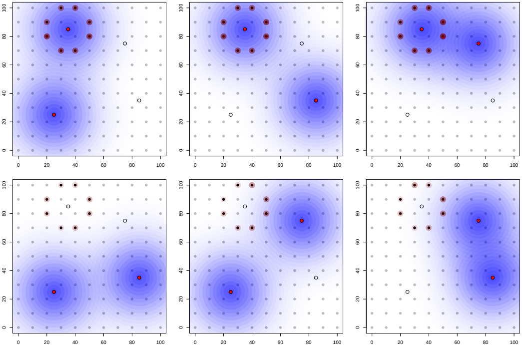
Defining effects



- Let $d_i(x)$ be distance between treatment node i and x.
- "Marginalized individualistic response" (MIR) for distance d:

$$\tau(d;\alpha) = \mathrm{E}_{\mathcal{N}}\left[\mathrm{E}_{\mathcal{X}}\left[\tau_{ix}(\alpha) | d_i(x) = d\right]\right]$$

Defining effects

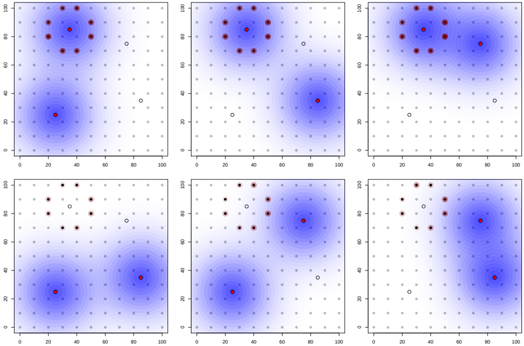


- Let $d_i(x)$ be distance between treatment node i and x.
- "Marginalized individualistic response" (MIR) for distance d:

$$\tau(d;\alpha) = \mathrm{E}_{\mathcal{N}}\left[\mathrm{E}_{\mathcal{X}}\left[\tau_{ix}(\alpha) | d_i(x) = d\right]\right]$$

On average, how are units at a specified distance from an intervention point affected by switching on treatment at that point, given effects emanating from other intervention points?

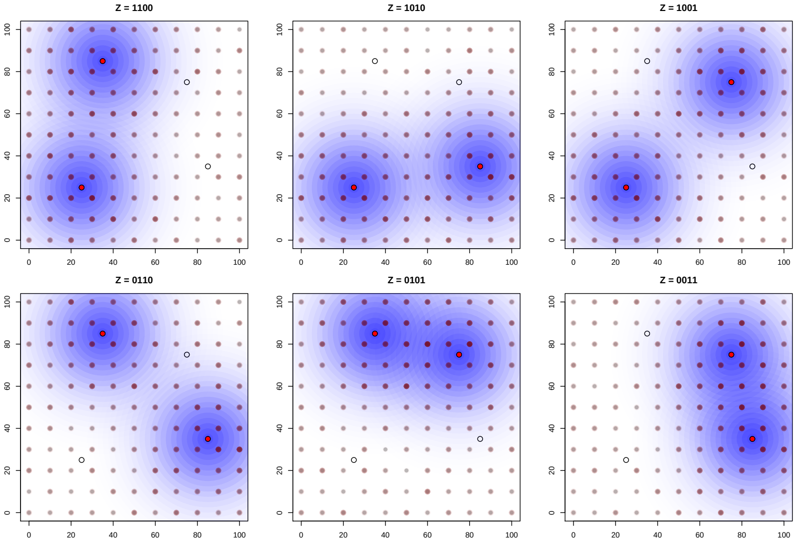
Defining effects



- MIR: $\tau(d;\alpha) = \mathrm{E}_{\mathcal{N}}\left[\mathrm{E}_{\mathcal{X}}\left[\tau_{ix}(\alpha) | d_i(x) = d\right]\right]$
- Nature of the effect depends on the randomization scheme, which defines the nature of “ambient exposure.”
- Potential outcomes at d combine effects of (i) how exposure varies in distance from the node and (ii) how the population varies in distance from the node.

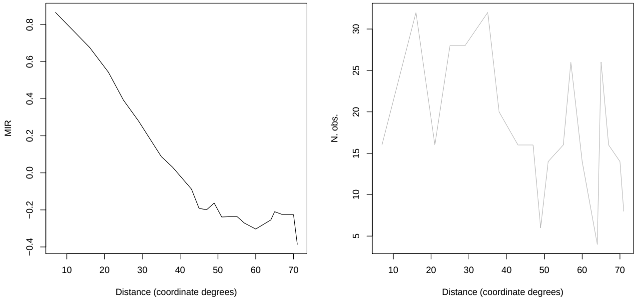
Defining effects

We can estimate the MIR for our toy example. Possible assignments:



Defining effects

The true MIRs given these assignments (left) and population measures for each d:



- Note that negative effects at a large distance.
- This is specific to *complete randomization* with a *small* N: if you are far from node i, switching on node i implies switching off a node that is *close*.
- Phenomenon disappears w/ Bernoulli randomization or N large.

Identification and estimation

- *Proposition 1*
  Define the following difference in integrated conditional means,

$$\tau(d;\alpha;\mathbf{Z}) = \mathrm{E}_{\mathcal{N}|Z_i=1}\left[\mathrm{E}_{\mathscr{X}}[Y_x|d_{ix}=d, Z_i=1]\right] - \mathrm{E}_{\mathcal{N}|Z_i=0}\left[\mathrm{E}_{\mathscr{X}}[Y_x|d_{ix}=d, Z_i=0]\right]. \qquad (1)$$

  Then, under the design, the difference in integrated conditional means is unbiased for $\tau(d;\alpha)$, as

$$\tau(d;\alpha) = \mathrm{E}_{\mathscr{Z}}[\tau(d;\alpha;\mathbf{Z})].$$

Identification and estimation

- *Proposition 1*
  Define the following difference in integrated conditional means,

$$\tau(d;\alpha;\mathbf{Z}) = \mathrm{E}_{\mathscr{N}|Z_i=1}\left[\mathrm{E}_{\mathscr{X}}[Y_x|d_{ix}=d, Z_i=1]\right] - \mathrm{E}_{\mathscr{N}|Z_i=0}\left[\mathrm{E}_{\mathscr{X}}[Y_x|d_{ix}=d, Z_i=0]\right]. \qquad (1)$$

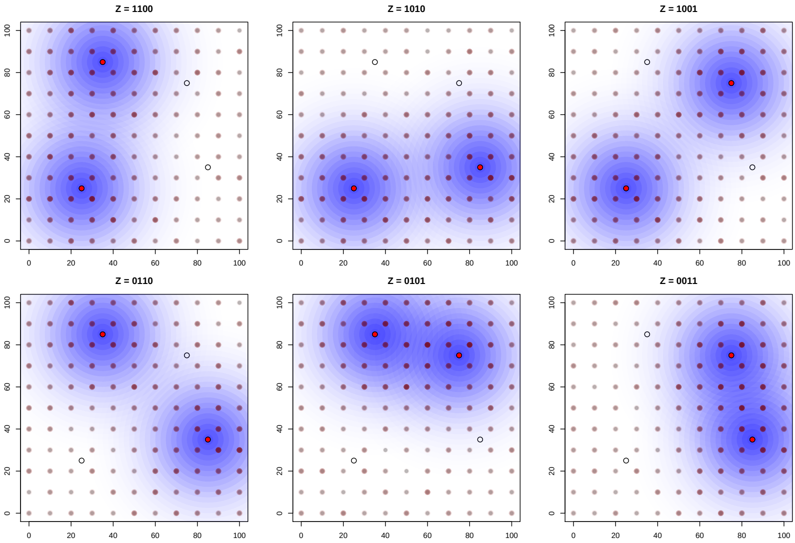
  Then, under the design, the difference in integrated conditional means is unbiased for $\tau(d;\alpha)$, as

$$\tau(d;\alpha) = \mathrm{E}_{\mathscr{Z}}[\tau(d;\alpha;\mathbf{Z})]. \qquad (2)$$

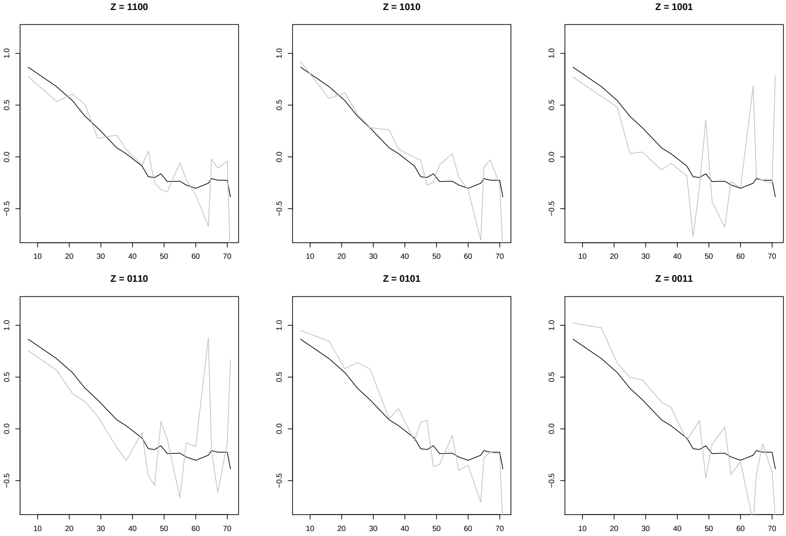
- Estimation with non-parametric analogue estimator:

$$\hat{\tau}(d) = \frac{1}{Np}\sum_{i=1}^{N} Z_i \bar{Y}_i(d) - \frac{1}{N(1-p)}\sum_{i=1}^{N}(1-Z_i)\bar{Y}_i(d). \qquad (3)$$

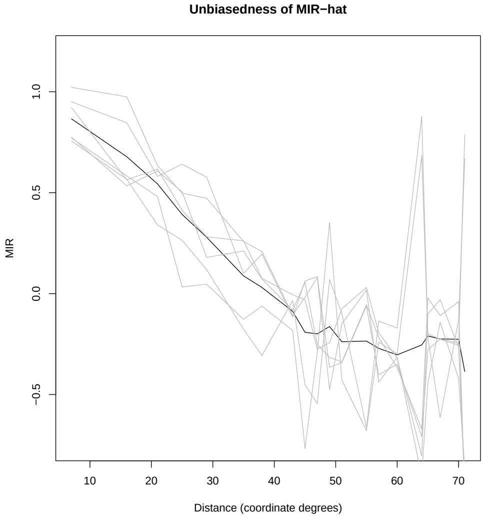
Identification and estimation



Identification and estimation



Identification and estimation



Identification and estimation

- Consistency for nonparametric estimator follows from local dependence of outcomes.
- *Proposition 2*:
  In addition to the design, suppose that

$$\sum_{i=1}^{N}\sum_{j\neq i}\mathrm{Cov}\left[Z_i\bar{Y}_i(d), Z_j\bar{Y}_j(d)\right] = o(N^2), \tag{4}$$

$$\sum_{i=1}^{N}\sum_{j\neq i}\mathrm{Cov}\left[\bar{Y}_i(d), \bar{Y}_j(d)\right] = o(N^2), \tag{5}$$

  and

$$\mathrm{E}_{\mathscr{Z}}[Z_i] = \bar{p} \text{ for all } i = 1, \ldots, N. \tag{6}$$

  Then as $N \to \infty$, $\hat{\tau}(d) \to \tau(d;\alpha)$.

Estimation and inference

Efficiency:

- Sometimes we don't observe outcomes over all x.
- Potential efficiency gains by using interpolations.
- We consider kriging-based interpolation:

$$Y_x = P(x) + Z(x) + e_x,$$

where $P(x)$ is a polynomial expansion of the x coordinates ("drift"), x, $Z(.)$ is a mean zero Gaussian field with a covariance function of arbitrary decay (adaptive kernel estimator), and $e(.)$ is independent mean-zero error with finite variance.

Estimation and inference

Efficiency:

- Sometimes we don't observe outcomes over all x.
- Potential efficiency gains by using interpolations.
- We consider kriging-based interpolation:

$$Y_x = P(x) + Z(x) + e_x,$$

  where $P(x)$ is a polynomial expansion of the x coordinates ("drift"), x, $Z(.)$ is a mean zero Gaussian field with a covariance function of arbitrary decay (adaptive kernel estimator), and $e(.)$ is independent mean-zero error with finite variance.
- *Proposition 3*:
  Given the design, denote the size of the sample of evaluation points as $|\mathscr{X}|$, and suppose that as $|\mathscr{X}| \to \infty$, $\hat{\bar{Y}}_i(d) \to \bar{Y}_i(d)$ for all i and d. Then, as both $N \to \infty$ and $|\mathscr{X}| \to \infty$, $\hat{\tau}_k(d) \to \tau(d;\alpha)$.

Estimation and inference

Inference:

- Fisher-style randomization inference.
- Permute treatment assignments under sharp null:

$$H_0 : Y_x(\mathbf{z}) = Y_x(\mathbf{z}') \text{ for all } \mathbf{z}, \mathbf{z}' \in \mathscr{Z}.$$

- Let $\hat{\tau}^0_{\mathbf{z}'}(d)$ be an estimate under H_0 and treatment permutation $\mathbf{z}'$.
- Exact Fisher p-value for a test at level α of the sharp null:

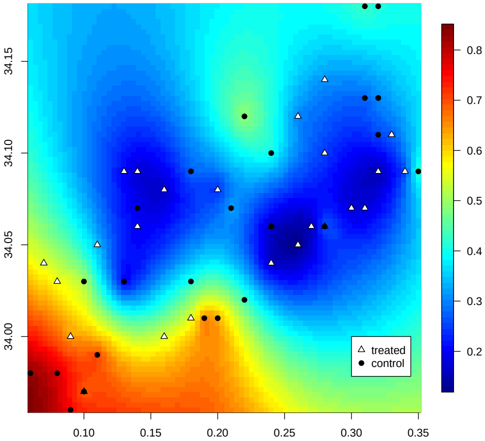
$$p = \Pr(\hat{\tau}^0_{\mathbf{Z}}(d) \leq \hat{\tau}(d) | H_0).$$

- In application we show a graphical approach.

Application

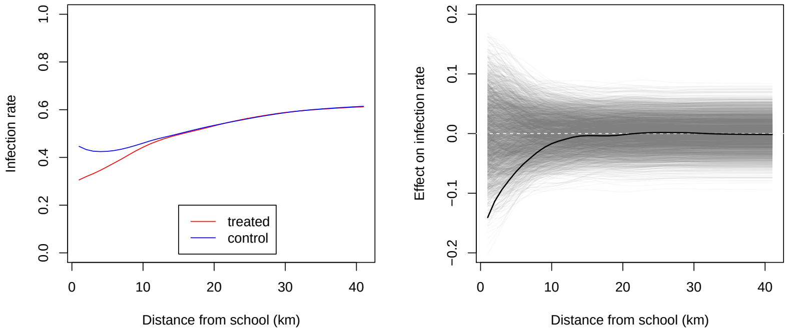
- Miguel and Kremer (2004) deworming RCT.
- Deworming treatments administered to kids at schools.
- Contagiousness of disease makes this a herd immunity process.
- We look at effect on worms infection rates (main causal mechanism).

Application



- Kriging fit to Miguel-Kremer data.
- Outcome is worms infection rate.
- (Note: Plot is rotated with northing on the *x*-axis.)

Application



- (left) Integrated means over distance from treated (red) and control (blue) locations
- (right) marginalized individualistic effects over distance (black), and sharp null distribution (gray)

Conclusion

- Towards spatial estimation and inference without SUTVA.
- Design-based methods avoid arbitrary assumptions in studying spatial spillover effects:
  - E.g., Miguel and Kremer estimated spillover effects by defining distance strata.
  - "Worm wars" controversy: results robust to other ways of characterizing spillover?
  - The methods here tell us what the design delivers with few additional assumptions.
- Estimand has policy relevance—characterizes effect associated with the "marginal" treated locality.