Treatment regimes and social networks

Mary E. Thompson, University of Waterloo

Time Series, Spatial Processes and Asymptotic Methods
Kelowna, BC
May 30, 2019

Reg Kulperger

- probability theory
- models of stochastic systems
- cellular automata and fire spread
- dynamics of disease transmission
- spatial statistical tools and genome-wide mutation thundershower detection
- ... a short distance to networks?

Joint work with Cong Jiang and Michael Wallace

Outline

Outline

Treatment Regime

- There is a desired outcome or outcome distribution, and a choice of treatments to try to bring it about.
- The best treatment for an individual at a certain stage could depend on the values of certain covariates, and the stochastic future, . . .
- . . . which would depend on the history of responses to treatments at earlier stages.
- A treatment regime is a rule for assigning treatment.
- A dynamic treatment regime is one that can adapt over time, such as by taking into account responses to treatments at earlier stages.

Relationship to Causal Inference

- To obtain a model for response to treatment, we:
  - conduct experiments under various conditions, OR
  - try to make inferences from observational studies
    - statistical learning; double robustness (e.g. Wallace and Moodie, 2015)
- Longitudinal data allow us to observe how history affects response to treatment.
- Estimation of model parameters allows us to compute optimal treatment regimes in new contexts.

SUTVA Assumption

Stable Unit Treatment Value Assumption

- An important assumption in causal inference.
- It includes the "no interference" assumption: "the observation on one unit should be unaffected by the particular assignment of treatments to the other units".
- It may be realistic for medical treatments, but not so much for behavioural interventions.
- When there *is* interference, it could benefit individuals and/or the group.

Networks

A network is a set of objects and their interconnections: “nodes” and “edges” or “links”.

e.g. Zhang et al (2011): nodes are genes and links are co-expression.

Social Networks

- nodes are people or groups of people
- links are “contact” or acquaintance
- Nodes are directly connected individuals:
  - friendship networks
  - collaboration networks
- Hierarchical networks:
  - individuals in couples or families
  - children in school classes
  - residents in nursing home wings

Treatment effects in networks

- An individual of primary interest is the "ego"
- Those to whom the ego is linked are the "alters"
- Treating the ego has a "direct effect" on the ego
- Treating an alter may have an "indirect effect" on the ego

Outline

Health behaviour in a family setting

- A couple who smoke cigarettes (ego, alter) and would like to quit
- Treatment is either a nicotine replacement therapy (NRT) or will-power (WP) alone. Four possible treatment combinations:

$$(NRT, NRT), (NRT, WP), (WP, NRT), (WP, WP)$$

Vaccine Efficacy

Ogburn and Vanderweele (2017)

- A couple of susceptibles (ego, alter) within a larger network; not necessarily isolated
- Effects of vaccinating the alter:
  - alter does not become infected, or
  - alter infectiousness is reduced
- Individual objective: minimize the probability of contracting the disease
- Population objective: most cost-effective vaccination-cum-isolation strategy to contain an epidemic

Model estimation

Case where ego is not vaccinated; exposed only to the alter

Basic GLM equations for one (ego,alter) pair

C: covariates

V_a: indicator for alter vaccination

Y_a^t: indicator for alter being infective at time t

Y_e^t: indicator for ego being infective at time t

s: incubation period

$$logit\{E[Y_e^{T+s} \mid V_a, Y_a^T = 1, C]\} = \gamma_0 + \gamma_1 V_a + \gamma_2' C,$$
$$logit\{E[Y_a^T \mid V_a, C]\} = \eta_0 + \eta_1 V_a + \eta_2' C.$$

Network persuasion campaign

Aronow and Samii (2017)

- 28 of 56 schools were randomly selected to host an anti-conflict program
- within every school, 40 - 64 students were non-randomly selected as *eligible* to participate
- within each host school half the eligibles were randomized to participate, with blocking on gender, grade and a measure of network closure
- friendship network data had been collected at the beginning of the year
- five treatment conditions
- outcome: willingness to endorse anti-conflict norms
- covariate: number of friends (network degree)
- indirect effects confirmed

Recruitment to game

Su et al (2018)

- $n = 961$ users of a social media platform and their friendship network
- treatment of a node ($a = 0$ or 1): one of two types of invitation
- $Y_i = -log(T_i)$ where $T_i =$ time until i joins the game, once invited
- $Y_i(a_i, s_i) = \mu(x_i, X_i) + \eta a_i + \gamma_2 \sum_j W_{ij} \eta a_j + a_i \theta' x_i + \gamma_3 \sum_j W_{ij} a_j \theta' x_j + \epsilon_i$
- optimal treatments to maximize $\frac{1}{n} \sum_i E(Y_i)$: $I\{[(1 + \gamma_2 \sum_{j \neq i} W_{ij})\eta + (1 + \gamma_3 \sum_{j \neq i} W_{ij})\theta' x_i] > 0\}$
- covariates are age, level of internet activity

Genetic networks

Emmert-Streib et al (2014); Cava et al (2018)

- genetic pathways, where genes tend to be co-expressed, are inferred from microarray data
- membership in one or more regulatory networks can help to identify "driver" genes
- there may be potential to "treat" a cancer driver gene, affecting its action and the action of others in its pathway

Outline

Network

- nodes or vertices are known or hypothesized
- “default” connections/edges are established, through family relationships, friendships, school class, workplace, correlation of expression
- a sample of edges may be known, e.g.limited number of connections each of a sample of nodes
- network may be hierarchical, with connections across clusters of connected units
- edges may be directed or undirected

Treatments and outcomes

Treatments:

- a medical treatment to change state of health or immunity
- an intervention to change behaviour to promote health
- an intervention to change "state" (being in or out of game; quality of being dangerous or not)

Outcomes:

- measure of (improved) health
- adoption of a behaviour
- a state indicator (being informed or not; being dangerous or not)

Outline

More general decision problems

Moving away from continuous outcomes and additivity:

- the problem of choosing treatments to maximize the expected good outcome for an individual is tractable
- it may not be possible to do this simultaneously for all
- the best outcome for the system as a whole is not necessarily best for each individual, especially when costs are limited
- e.g. maximizing the number of students willing to endorse norms for the least program cost

Maintaining double robustness

In an observational study, how do we

- separate the treatment-free mean response from the dependence of treatment effects on covariates?
- in estimation, effectively balance the covariates so as to obtain a true characterization of the treatment effect when either the treatment propensity model or the treatment-free model for the outcome is incorrect?

Asymptotic theory

Leung (2019)

- $\mu_h(d, t, \gamma) = E[h(Y(d, t, \gamma, \epsilon_1))]$
- d is treatment of the individual
- t is number of treated neighbours
- γ is number of neighbours (degree - fixed)
- E is with respect to the distribution of ϵ_1
- interest centres on direct effect: $\mu_h(d, t, \gamma) - \mu_h(d', t, \gamma)$
- …and indirect effects: $\mu_h(d, t, \gamma) - \mu_h(d, t', \gamma)$

Sparsity conditions on the expanding network permit consistency and asymptotic normality results.

Happy retirement, Reg!

References

- Aronow, P. M. and Samii, C. (2017) Estimating average causal effects under general interference, with application to a social network experiment *Annals of Applied Statistics* 11, 1912-1947.
- Cai, J., De Janvry, A. and Sadoulet, E. (2015) Social networks and the decision to insure. *American Economic Journal: Applied Economics* 7, 81-108.
- Cava, C., Bertoli, G., Colaprico, A., Olsen, C., Bontempi, G. and Castiglioni, I. (2018) Integration of multiple networks and pathways identifies cancer driver genes in pan-cancer analysis. *BMC Genomics* 19, 25.
- Emmert-Streib, F., Dehmer, M. and Haibe-Kains, D. (2014) Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Frontiers in Cell and Developmental Biology*, 19 August 2014.
- Leung, M. P. (2019) Treatment and spillover effects under network interference. Manuscript.
- Miguel, E. and Kremer, M. (2004) Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72, 159-217.
- Ogburn, E. L. and VanderWeele, T. J. (2014) Causal diagrams for interference. *Statistical Science* 29, 559-578.

References

- Ogburn, E. L. and VanderWeele, T. J. (2017) Vaccines, contagion and social networks. *The Annals of Applied Statistics* 11, 919-948.
- Ramadan, E., Alinsaif, S. and Hassan, M. R. (2016) Network topology measures for identifying disease-gene association in breast cancer. *BMC Bioinformatics* 17, 274.
- Su, L., Lu, W. and Song, R. (2018) Modelling and estimation for optimal treatment decision with interference. *Wiley*. DOI: 10.1002/sta4.219
- Sussman, D. L. and Airoldi, E. M. (2017) Elements of estimation theory for causal effects in the presence of network interference. Manuscript.
- Wallace, M. P. and Moodie, E. E. M. (2015) Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* 71, 636-644.
- Zhang, J., Yang, Y. et al (2011) Identification of hub genes related to the recovery phase of irradiation injury by microarray and integrated gene network analysis. *PloS one* 6(9): e24680.